Situation-Aware Cyber-Physical-Social System for Cultural Heritage Protection

Rosario Gaeta

UNIVERSITY OF SALERNO

DEPARTMENT OF INDUSTRIAL ENGINEERING

Ph.D. Course in Industrial Engineering
Curriculum in Electronic Engineering
XXXIII Cycle

Situation-Aware Cyber-Physical-Social System for Cultural Heritage Protection

Supervisor
Chiar.mo Prof. Massimo De Santo

Ph.D. Student
Rosario Gaeta

Scientific Referees
Chiar.mo Prof. Marco Carratù
Chiar.mo Prof. Gian Domenico Licciardo

Ph.D. Course Coordinator
Chiar.mo Prof. Massimo De Santo

In loving memory of my brother Mario.

Abstract

Cultural heritage critical infrastructures (CHCIs), including archaeological sites, museums, and historic structures, are vital to a country's identity and economy. Safeguarding them is an ethical imperative that also supports employment and territorial attractiveness. Yet their preservation is threatened by aging materials, extreme weather and other environmental phenomena, and visitors who may cause damage, unintentionally or maliciously. CHCIs therefore require regular inspection and maintenance. These procedures are traditionally performed by humans and are often slow, error-prone, and resource-intensive. For example, if the roof of a covered structure containing priceless frescoes is damaged, prompt action is crucial; otherwise, a single storm or prolonged rainfall could cause severe loss.

At the same time, in CHCI maintenance not everything can be automated. Final decisions must always be made by human operators, since any intervention must be justified and priorities in maintenance plans must be assessed by experts. A Human-Machine methodological approach is therefore essential: the user retains final responsibility, while their Situation Awareness is actively supported and maintained at a high level through new technologies that reduce workload and optimize task performance.

We propose to see the overall system, comprising the CHCI, the operators and visitors who interact with it, and the supporting technologies, as a Cyber-Physical-Social System (CPSS). In particular, we propose a novel Situation-Aware CPSS methodological approach for cultural heritage protection that can be adapted and replicated by public and private actors. Its implementation relies on sensors and data-acquisition paradigms, Artificial Intelligence (AI) to automate and accelerate the recognition of risk situations and maintenance issues, and adaptive interfaces capable of presenting the information gathered.

Our research begins with a state-of-the-art review of AI for cultural heritage preservation to identify the most modern and effective technologies to integrate into the model. To recognize risk, the situation must also be modeled. Building on the literature, we introduce a goal-driven Cultural Heritage Vulnerability Index (CHVI), computed using Context Space Theory (CST), to assess the vulnerability of areas and structures based on identified maintenance issues

and environmental conditions.

The work also incorporates and evaluates an innovative data imputation technique based on association rules and granular computing to reconstruct missing data and ensure a more accurate perception of the CHCI environment. The approach is tested on real-world IoT datasets established in the literature. Automatic, rapid, and effective recognition of CHCI maintenance issues is further achieved through a novel sustainable method for identifying damage and weedy vegetation on the roofs of archaeological structures, evaluated on datasets related to the Pompeii archaeological site.

All these elements converge in VERGIL, a CPSS platform designed to support human operators at Pompeii. VERGIL includes an adaptive interface with a novel Oscillatory Goal Suggestion approach that, in emergency situations, redirects the user's attention to elements requiring urgent action, while avoiding unnecessary distraction when the situation does not require it. Finally, VERGIL is evaluated using the Situation Awareness Global Assessment Technique (SAGAT) across multiple scenarios set at Pompeii, testing with human operators whether the system effectively supports their Situation Awareness and decision-making abilities.

Acknowledgment

Foreign readers will forgive me for writing my acknowledgments in Italian, but I believe some things should be said in one's native language.

Un grazie speciale va al mio Supervisor, il Prof. *Massimo De Santo*, che fin dall'inizio mi ha supportato e guidato nella mia esperienza di dottorato. Ti ringrazio anche per esserci sempre stato, soprattutto nei momenti più complicati, quando il tuo aiuto era necessario: sei sempre stato lì per me.

Ringrazio molto il Prof. *Francesco Colace*, che è stato per me un grande motivatore e un insegnante in questi anni. Ti ringrazio moltissimo anche per avermi dato il privilegio di lavorare nella realtà complessa e prestigiosa del Parco Archeologico di Pompei. Le problematiche che abbiamo affrontato lungo questo percorso mi hanno fatto crescere come persona e come ricercatore e porterò questi insegnamenti sempre con me, insieme ai momenti che abbiamo condiviso, sia dentro che fuori dall'ambito lavorativo.

Un grande ringraziamento va al Prof. *Giuseppe D'Aniello*, che è mio insegnante dai tempi della triennale e che sopporta le mie pare da altrettanto tempo. Ti ringrazio non solo perché ci sei sempre, ogni volta che ho bisogno di te, ma perché sei una di quelle poche persone che riescono ad aprirti la mente e a insegnarti qualcosa anche solo con una chiacchierata. Ma, più di tutto, ti ringrazio perché riesci a tirare fuori il meglio da me anche quando non lo vedo nemmeno io.

Ringrazio tanto la Prof.ssa *Stefania Tomasiello*, con cui ho avuto il piacere di collaborare, per i preziosi insegnamenti che mi ha trasmesso e continua a trasmettermi, contribuendo in modo significativo alla mia crescita come ricercatore.

Un sentito ringraziamento va anche al Prof. *Francesco Flammini*, che è stato la mia guida durante il mio periodo in Svizzera al SUPSI. Lo ringrazio infinitamente per avermi accolto e per avermi dato l'opportunità di lavorare insieme a lui e al suo gruppo di lavoro. Ringrazio di cuore tutti gli amici di Lugano con cui ho condiviso tante giornate di spensieratezza. In particolare, voglio ringraziare *Franca* per essere stata la prima ad accogliermi e a farmi da guida in un ambiente per me nuovo, ma anche *Giorgia* e *Lorenzo*, che mi hanno fatto sentire a casa, tra chiacchierate durante le pause, serate a giocare a

Mario Kart e pizze fatte in casa, tra le ultime prima di diventare, mio malgrado, celiaco.

Un grande ringraziamento va ai valutatori di questa tesi per il loro prezioso lavoro.

Ringrazio *Netcom Group* per aver cofinanziato la mia borsa e per avermi dato, quindi, l'opportunità di intraprendere questo percorso, e *Alfredo*, mio punto di riferimento in azienda, dal quale nei momenti di confronto imparo sempre qualcosa.

Ringrazio molto *Fabio*: è un piacere collaborare con te e sono orgoglioso dei risultati che abbiamo raggiunto insieme.

Ringrazio il Centro ICT per i Beni Culturali e tutti i suoi componenti, in particolare: *Angela, Angelo, Annabella, Annachiara, Carmine, Connie, Domenico, Flavia, Giusy, Liliana, Marco, Mario, Michele, Milinda, Nadeeka, Rocco e Rosario*. Mentre scrivo queste righe mi rendo conto che, trascorrendo la maggior parte del tempo al lavoro, siete le persone con cui ho condiviso più tempo durante questi anni. Vi dico grazie per tutti i momenti di serenità ma anche per quelli di lavoro, per i nostri pranzi e anche per gli insegnamenti che ognuno di voi mi ha dato. Ringrazio anche tutte le persone con cui ho collaborato in questi anni. Ringrazio in maniera speciale *Carmine*, che è stato il mio compagno di dottorato fin dall'inizio. Nelle nostre innumerevoli pause c'è sempre stato spazio per il confronto lavorativo, ma anche e soprattutto per parlare della vita. Ti ringrazio per tutta la saggezza che mi hai trasmesso in tante occasioni. Un grande grazie anche a *Milinda*: dai nostri confronti esco sempre arricchito da tanti insegnamenti lavorativi e dai racconti delle tue esperienze di vita.

Ringrazio *Nicola*, mio collega di dottorato e vice-chair nell'IEEE Student Branch. Durante il nostro percorso non sono mai mancate le occasioni per confrontarci e aiutarci a superare le sfide e le difficoltà che si presentavano, sia con il dottorato sia con il Branch.

Un grande grazie lo voglio dire a *Federica* e *Isabella*, le (ex) supertesiste del Centro ICT, perché durante il periodo in cui siete state con noi in laboratorio avete portato tanta gioia e anche un po' di ansia, anche se quella, in realtà, c'era già a prescindere. Ricorderò sempre con grande affetto i momenti che abbiamo passato insieme.

Voglio dedicare qualche parola anche ai ragazzi del laboratorio KnowMIS: *Alessandro, Camilla, Mario e Roberto*. Ci siamo conosciuti da poco, ma sono molto felice perché sta nascendo un bel gruppo e spero che raggiungeremo tanti traguardi insieme.

Il grazie più grande va alla mia famiglia, a cui voglio bene più di qualsiasi altra cosa.

Grazie a mio *padre* per esserci sempre per me quando ho bisogno, anche quando sono troppo orgoglioso per chiedere aiuto. Tutti i traguardi che ho

raggiunto sono dovuti anche al perfezionismo che mi hai trasmesso e che guida tutto ciò che faccio. Sono certo che tu sappia che, nel seguire la strada che sto percorrendo, spero sempre di renderti orgoglioso.

Grazie a mio fratello *Vito*, a cui non sarò mai in grado di restituire tutto quello che ha fatto per me. Tutti i traguardi che ho raggiunto sono stati possibili anche grazie a te, perché il tuo esserci sempre mi ha aiutato in tutti i momenti difficili che abbiamo passato e che continuiamo a passare insieme. Sarai sempre il mio rifugio sicuro e spero di esserlo anche io per te.

Grazie alla mia cara *mamma*, che da sempre è il mio modello e il mio punto di riferimento. Hai vissuto senza mai scegliere la strada più facile, cercando di essere sempre corretta con tutti e senza aspettarti nulla in cambio. Ti sei dedicata a noi figli mettendo te stessa in secondo piano, e mi dispiace che la vita non ti abbia dato tutto quello che meritavi. Io non sono mai stato all'altezza di tutto l'amore che mi hai donato, e mi dispiace non riuscire, purtroppo, a starti vicino e ad aiutarti come vorrei quando ne hai bisogno. Voglio però che tu sappia che ogni cosa buona che riesco a fare è dedicata a te. Sei la mamma migliore che si possa desiderare.

Grazie a mio fratello *Mario*, a cui questa tesi è dedicata. Anche se non ci sei dall'inizio del mio dottorato, spero tu possa vedere tutto questo e che il mio percorso ti renda orgoglioso di me come io lo ero di te.

List of Publications

Parts of this thesis are based on the following publications. As a coauthor, I was involved actively in the research, planning and writing these papers.

Journal Publications

i. F. Clarizia, M. De Santo, R. Gaeta, and R. Loffredo, "Enhancement Large Language Models Domain through Ontology-Based Retrieval-Augmented Generation," International Journal on Semantic Web and Information Systems, IGI Global, 2025, doi: 10.4018/IJSWIS.392507.

ii. T. M. H. Gedara, R. Gaeta, S. Tomasiello, and R. El Shawi, "Towards Explainable Multimodal Fake News Detection: A Fusion of Prompt Learning and Automated Machine Learning," IEEE Access, 2025, IEEE, doi: 10.1109/ACCESS.2025.3627504.

iii. M. Casillo, F. Colace, R. Gaeta, A. Lorusso, and M. Pellegrino, "Artificial Intelligence in Archaeological Site Conservation: Trends, Challenges, and Future Directions," Journal of Computer Applications in Archaeology, 2025, Ubiquity Press, doi: 10.5334/jcaa.207.

iv. F. Colace, R. Gaeta, A. Lorusso, M. Pellegrino, and D. Santaniello, "New AI Challenges for Cultural Heritage Protection: A General Overview," Journal of Cultural Heritage, 2025, Elsevier, doi: 10.1016/j.culher.2025.07.019.

v. F. Colace, G. D'Aniello, M. De Santo, R. Gaeta, and G. Zuchtriegel, "Situation-aware Cyber–Physical–Social System for Cultural Heritage," Intelligent Systems with Applications, Elsevier, 2025, doi: 10.1016/j.iswa.2025.200544.

vi. F. Colace, M. De Santo, R. Gaeta, and R. Loffredo, "Detection of Maintenance Issues from UAV Images of Archaeological Sites: A YOLO-

Based Tool," Multimedia Tools and Applications, 2025, Springer, doi: 10.1007/s11042-025-20773-7.

vii. E. Bellini, G. D'Aniello, F. Flammini, and R. Gaeta, "Situation Awareness for Cyber Resilience: A Review," International Journal of Critical Infrastructure Protection, 2025, Elsevier, doi: 10.1016/j.ijcip.2025.100755.

viii. M. Casillo, F. Colace, R. Gaeta, A. Lorusso, D. Santaniello, and C. Valentino, "Revolutionizing Cultural Heritage Preservation: An Innovative IoT-Based Framework for Protecting Historical Buildings," Evolutionary Intelligence, Springer, 2024, doi: 10.1007/s12065-024-00959-y.

ix. R. Mosca, M. De Santo, and R. Gaeta, "Ontology Learning from Relational Database: A Review," Journal of Ambient Intelligence and Humanized Computing, 2023, Springer, doi: 10.1007/s12652-023-04693-8.

Conference Publications

i. R. Gaeta, F. Corradini, M. De Santo, F. Flammini, and H. Ge, "Improving Situation Awareness and Self-Adaptation in Autonomous Wheelchair-Drone Systems through Floor Surface Anomaly Detection," in 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2025), IEEE. doi: 10.1109/SMC58881.2025.11342578

ii. R. Gaeta, G. D'Aniello, and V. Zampoli, "Situation-aware Adaptive Interfaces for Cultural Heritage Based on Oscillatory Attention Dynamics," in 2025 IEEE International Conference on Cyber Humanities (CH 2025), IEEE. (Under Publication)

iii. M. Casillo, F. Colace, R. Gaeta, M. Lombardi, D. Santaniello, and C. Valentino, "An Architecture Based on Recommender Systems and Context Awareness to Suggest Personalized Cultural Experiences," in Tenth International Congress on Information and Communication Technology (ICICT 2025), Springer, doi: 10.1007/978-981-96-6932-5_18.

iv. L. Cecere, M. C. De Simone, S. P. Dembele, R. Gaeta, A. Lorusso, and D. Santaniello, "New Technologies for Concrete Bridge Monitoring: An Overview," in Tenth International Congress on Information and Communication Technology (ICICT 2025), Springer, doi: 10.1007/978-981-96-6932-5_15.

v. M. De Santo, F. Flammini, R. Gaeta, H. Ge, and Z. U. Rehman, "An AutoML Approach for the Efficient Classification of Damaged Roofs Using Fuzzy-Transform," in 2025 IEEE 5th International Conference on Human-Machine Systems (ICHMS 2025), IEEE, doi: 10.1109/ICHMS65439.2025.11154372.

vi. L. Aliberti, G. D'Aniello, M. De Santo, and R. Gaeta, "A Situation-Aware Cloud-Edge Architecture for Cultural Heritage Management," in 2025 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA 2025), IEEE, doi: 10.1109/CogSIMA64436.2025.11079541.

vii. R. Frontino, R. Gaeta, and R. Mosca, "Mining Knowledge from Data: The Case of Ontology Learning," in The Italian Workshop on Neural Networks (WIRN 2023), 2025, Springer, doi: 10.1007/978-981-96-0994-9_30.

viii. F. Colace, S. P. Dembele, R. Gaeta, A. Lorusso, D. Santaniello, and C. Valentino, "Enhancing Agricultural Practices Through IoT and Decision Tree Analytics: A Case Study on Autonomous Irrigation Management," in World Conference on Information Communication Systems, Software, Security and Sustainability (WorldS4 2025), Springer, doi: 10.1007/978-981-97-9559-8_20.

ix. F. Colace, R. Gaeta, D. Santaniello, A. Troiano, and C. Valentino, "A Framework Based on Internet of Things and Recommender Systems for Thermal Facilities," in Ninth International Congress on Information and Communication Technology (ICICT 2024), Springer, doi: 10.1007/978-981-97-5035-1_8.

x. M. Casillo, R. Gaeta, A. Lorusso, F. Marongiu, D. Santaniello, and C. Valentino, "A Novel Architecture for Enhancing Museum Visits Through Recommender Systems, Digital Storytelling, and NFT," in Ninth International Congress on Information and Communication Technology (ICICT 2024), Springer, doi: 10.1007/978-981-97-5035-1_9.

xi. F. Colace, R. Gaeta, M. Lombardi, A. Lorusso, D. Santaniello, and C. Valentino, "An Architecture for Cultural Heritage Enhancement Based on the Internet of Things and a Hybrid Context-Aware Recommender System," in 2024 IEEE International Conference on Automation/26th Congress of the Chilean Association of Automatic Control (IEEE ICA-ACCA 2024), IEEE, doi: 10.1109/ICA-ACCA62622.2024.10766774.

xii. F. Colace, R. Gaeta, A. Lorusso, and D. Santaniello, "Smart Restoration: AI for Historical Façade," in 2024 IEEE Workshop on Complexity in Engineering (COMPENG 2024), IEEE, doi: 10.1109/COMPENG60905.2024.10741394.

xiii. F. Colace, M. De Santo, R. Gaeta, R. Loffredo, and L. Petti, "FAUNO: A Machine Learning-Based Methodology for Monitoring and Predictive Maintenance of Structures in Archaeological Parks Through Image Analysis," in 26th International Conference on Human-Computer Interaction (HCII 2024), Springer, doi: 10.1007/978-3-031-60611-3_24.

xiv. F. Clarizia, M. De Santo, R. Gaeta, and R. Mosca, "Method for Ontology Learning from an RDB: Application to the Domain of Cultural Heritage," in Pattern Recognition for Cultural Heritage, 5th International Workshop (PatReCH 2023 – ICIAP 2023), Springer, doi: 10.1007/978-3-031-51026-7_35.

Contents

Abstract iv

Acknowledgment vii

List of Publications xii

List of Figures xviii

List of Tables xxiii

1 Introduction 1
- 1.1 Challenges and Objective . . . 5
- 1.2 Approach . . . 9
- 1.3 Contributions . . . 11
- 1.4 Thesis Outline . . . 13

2 Theoretical Background 15
- 2.1 Predictive Maintenance . . . 16
  - 2.1.1 Maintenance in Cultural Heritage . . . 17
    - 2.1.1.1 Cultural Heritage Maintenance Lifecycle . . . 17
    - 2.1.1.2 Cultural Heritage Vulnerability Index . . . 20
- 2.2 Cyber-Physical-Social Systems and Digital-Twins . . . 21
- 2.3 Situation Awareness . . . 22
  - 2.3.1 Cognitive and Human Factors . . . 24
  - 2.3.2 Systems Design Factors . . . 25
  - 2.3.3 SA Demons . . . 26
- 2.4 Summary . . . 27

3 AI Challenges for Cultural Heritage Protection 29
- 3.1 Methodology . . . 30
- 3.2 Trend Analysis . . . 32
- 3.3 Technical Analysis . . . 34

3.3.1 Movable Heritage 36
3.3.1.1 Overview and Key Findings 38
3.3.2 Immovable Heritage 39
3.3.2.1 Archaeological Sites 41
3.3.2.2 Historical Buildings 43
3.3.2.3 Urban Areas 47
3.3.2.4 Infrastructures 48
3.3.2.5 Museums 48
3.3.2.6 Synthesis and Key Findings 48
3.4 Discussion 51
3.5 Summary 53

4 Situation-aware Cyber-Physical-Social System for Cultural Heritage 55
4.1 Related Works 57
4.1.1 Cultural Heritage Maintenance Approach 57
4.1.2 SA for Cultural Heritage 58
4.2 Situation-aware CPSS 58
4.2.1 Physical-Layer 59
4.2.2 Social-Layer 60
4.2.3 Cyber-Layer 61
4.2.3.1 L0 Sensing 63
4.2.3.2 L1 Perception 64
4.2.3.3 L2 Comprehension 65
4.2.3.4 L3 Projection 66
4.2.3.5 CPSS mapping to CH Lifecycle 67
4.3 Cultural Heritage Vulnerability Index and Situation Modeling . 67
4.4 Proof of Concept 71
4.4.1 Cultural Heritage Vulnerability Index Calculation . . . 72
4.4.1.1 Static CHVI Contribution 72
4.4.1.2 Dynamic CHVI Contribution 74
4.4.2 Case Study 1: Archaeological Site of Pompeii 78
4.4.3 Case Study 2: Archaeological Site of Paestum 85
4.4.4 Case Study 3: Archaeological Site of the Colosseum . 87
4.4.5 Discussion 88
4.5 Summary 90

5 Sensors and Sensing 91
5.1 Acquisition Sensing 92
5.1.1 Aerial Images 93
5.1.2 Contextual Data 93
5.2 Offline Sensing 94

5.3 Online Sensing . . . 96
5.3.1 Related Works . . . 97
5.3.1.1 Traditional imputation techniques . . . 98
5.3.1.2 Association-rule Imputation . . . 98
5.3.1.3 Neural network-based Imputation . . . 99
5.3.1.4 Granular imputation . . . 99
5.3.2 GARDA: Granular Association-rule-based Data Imputation Approach . . . 100
5.3.2.1 Formal definition of the approach . . . 102
5.3.2.2 Granulation approach . . . 104
5.3.2.3 GrC Association-rule-based Imputation . . . 106
5.3.2.4 Value reconstruction and Traditional Imputation . . . 109
5.3.3 Evaluation . . . 110
5.3.3.1 Datasets . . . 110
5.3.3.2 Evaluation Protocol . . . 112
5.3.3.3 Results . . . 113
5.3.3.4 Discussion . . . 120
5.4 Summary . . . 123

6 Perception and Comprehension of Maintenance Issues 125
6.1 Background and Related Works . . . 128
6.1.1 Detection of Maintenance Issues . . . 128
6.1.1.1 Weedy Vegetation . . . 129
6.1.1.2 Damaged Conduits . . . 129
6.1.1.3 Broken Tiles . . . 130
6.1.1.4 Damaged Structures . . . 131
6.1.2 Classification of Damaged Roofs . . . 132
6.1.3 Sustainable AI . . . 133
6.1.4 AutoML . . . 134
6.2 FAUNO: YOLO Detection Approach . . . 135
6.2.1 FAUNO Approach . . . 135
6.2.2 FAUNO Detector Tool . . . 137
6.2.3 Evaluation . . . 140
6.2.3.1 Dataset . . . 140
6.2.3.2 Evaluation Protocol . . . 144
6.2.3.3 Results . . . 146
6.2.3.4 Discussion . . . 148
6.3 HydraML: a Hybrid Damaged Roofs AutoML Classifier for Cultural Heritage . . . 150
6.3.1 Methodology . . . 151
6.3.1.1 AutoML Approach . . . 152

6.3.1.2 Fuzzy-Transform Compression 153
6.3.2 Evaluation 155
6.3.2.1 Dataset 155
6.3.2.2 Evaluation Protocol 156
6.3.2.3 Results 158
6.3.2.4 Discussion 169
6.4 Summary 171

7 VERGIL: A SA Oriented Platform for Archaeological Site Protection 173
7.1 Background 174
7.1.1 Goal-Directed Task Analysis 175
7.1.2 Situation Awareness Measurement 176
7.1.2.1 Situation Awareness Global Assessment Technique 178
7.2 Pompeii Archaeological Site Protection GDTA 179
7.3 VERGIL Adaptive Interface 184
7.3.1 Main Components and Implementation 185
7.3.2 VERGIL Knowledge Base 187
7.3.2.1 JSON2OWL tool 189
7.3.2.2 Pompeii Archaeological Site Ontology . . . 191
7.3.3 VERGIL Adaptive Interface 192
7.3.3.1 Interface Modes 193
7.3.3.2 Oscillatory Goal Suggestion Approach . . . 194
7.3.3.3 VERGIL Goal Suggestion Implementation . 197
7.3.4 CHVIs Implementation 201
7.3.4.1 Contribution Functions 201
7.3.4.2 Goal-driven CHVIs 205
7.4 Evaluation 206
7.4.1 Evaluation Procedure 207
7.4.2 Evaluation Scenarios 210
7.4.2.1 Scenario 1 211
7.4.2.2 Scenario 2 212
7.4.2.3 Scenario 3 213
7.4.2.4 Scenario 4 214
7.4.2.5 Scenario 5 215
7.4.3 Statistical Significance Evaluation Procedure 218
7.4.3.1 Dependent Variables 219
7.4.3.2 Descriptive Statistics 219
7.4.3.3 Hypothesis Testing 219
7.4.3.4 OLS ANOVA 220
7.4.3.5 Scenario-specific Comparisons 221

7.4.3.6 Effect Sizes 221
7.4.4 Results and Discussion 222
7.4.4.1 Global Results 222
7.4.4.2 In-depth Scenario Analysis 226
7.5 Summary 232

8 Conclusions and Future Works **235**
8.1 Summary 235
8.2 Final Remarks 237
8.3 Limitations 240
8.4 Future Works 241

Bibliography **242**

List of Figures

1.1 Research Approach 12

2.1 Cultural Heritage Maintenance Lifecycle 19
2.2 Endsley Model for Situation Awareness 24

3.1 Survey Procedure 31
3.2 Bibliographic Coupling Network 33
3.3 Citation Sources Density Map 34
3.4 Authors Keywords Density Graph 35
3.5 General Taxonomy of Heritage and Assets 36
3.6 Classification of Movable Heritage and Type of Data 37
3.7 Classification of Movable Heritage and Goals 37
3.8 Immovable Heritage Classification with Type of Data 40
3.9 Immovable Heritage Classification with Goals 40

4.1 Situation-aware Cyber-Physical-Social System for Cultural Heritage Protection 63
4.2 Example of 3D contextual space. 69
4.3 Images of *Casa del Fauno* and *Casa del Labirinto*. 79
4.4 Images of *Casa di Gavius Rufus* and *Casa di Nave Europa*. . . 80
4.5 Damaged Conduits on *Casa del Fauno* and *Casa del Labirinto*, (Red rectangles Highlight the Damage). 81
4.6 Broken Tiles of *Casa di Gavius Rufus* and *Casa di Nave Europa* (Red Rectangles Highlight the Damage). 82
4.7 Barplot of the Standalone Static CHVI Contribution in July 2023 83
4.8 Barplot of the Pompeii CHVI Structures Across the Months of 2023 84
4.9 Satellite Images of Paestum 2024 (Arcgis, Google Earth Pro). . 85
4.10 Barplot of the Paestum CHVI Structures in 2024 86
4.11 Paestum Weedy Vegetation May 2024. 87
4.12 Satellite Images of *Foro Romano* 2023 (Google Earth). 87
4.13 *Foro Romano, Rome* Weedy Vegetation Apr 2023. 89

4.14 Barplot of the Colosseum CHVI Structures in 2023 90

5.1 SA-CPSS with Sensors and Sensing Details 92
5.2 Orthophoto of the Pompeii Site 95
5.3 Orthophoto of the Regio I 96
5.4 GARDA Architecture 101
5.5 GARDA Flow Diagram 103
5.6 Application of the Granulation Approach to an Incomplete Dataset. 105
5.7 Dataset Reconstructed with the Mean of the Intervals. The Initial Missing Values Reconstructed by the Algorithm are in Bold. 109
5.8 Correlation Matrix of the Ten Sensors Considered in the Intel Lab 111
5.9 Correlation Matrix of the 4 Sensors from the Moderate Correlation Version of the AirQuality Beijing Dataset 111
5.10 Evaluation Approach of GARDA on Various Datasets 114
5.11 Comparison Between the Real and Imputed Values for Four Variables of the Intel Lab Dataset at 50% Missing Rate under the Consecutive Error Model. 121

6.1 SA-CPSS with Perception and Comprehension Details 126
6.2 FAUNO Approach 136
6.3 FAUNO Design Architecture 138
6.4 Maintenance Issues Examples 142
6.5 FAUNO Labelling Format Conversion 143
6.6 FAUNO Visualization of Intersection Over Union (IoU) 145
6.7 Bar Plot of $mAP_{0.5}$ Results 147
6.8 Examples of Detection of Maintenance Issues in Images . . . 149
6.9 HydraML: Hybrid Damaged Roofs AutoML Classifier for Cultural Heritage 151
6.10 Tiles Roofs 156
6.11 Flat Roofs 156
6.12 PompeiiTile M-Autogluon and HydraML Cross-Validation Accuracy and F1 over Time of Training. 161
6.13 PompeiiTile M-Autogluon and HydraML Cross-Validation CPU and GPU over Time of Training. 162
6.14 PompeiiFlat M-Autogluon and HydraML Cross-Validation Accuracy and F1 over Time of Training. 163
6.15 PompeiiFlat M-Autogluon and HydraML Cross-Validation CPU and GPU over Time of Training. 164
6.16 PompeiiTile Cross Validation Confusion Matrix (%) 165

6.17 PompeiiFlat Cross Validation Confusion Matrix (%) 166
6.18 Tile Roofs Issues Severity. 167
6.19 Flat Roofs Issues Severity. 168
6.20 Tiles Roofs Mistakes Examples (True - Predicted) 170
6.21 Flat Roofs Mistakes Examples (True - Predicted) 170

7.1 SA-CPSS with Projection and Application Details 175
7.2 Macro Goal Hierarchy of Goal-Directed Task Analysis 180
7.3 1.0 Hierarchy Goal-Directed Task Analysis 181
7.4 1.1 Goals, Decisions and SA Requirements 182
7.5 1.2 Goals, Decisions and SA Requirements 183
7.6 1.3 Goals, Decisions and SA Requirements 184
7.7 Overview Map of the SA Interface 186
7.8 Contextual Data on the Interface 187
7.9 SA-CPSS with Knowledge Base Details 188
7.10 JSON2OWL Tool Interface: Page 1 190
7.11 Pompeii Archaeological Site Ontology Graph 192
7.12 Vulnerability Info Hovering the Label of an Insula 193
7.13 Label and Position of the Structures 194
7.14 Flats Map of the SA Interface 194
7.15 Alert States Model 196
7.16 Yellow Alert for Rainfall 198
7.17 Orange Alert for Rainfall 198
7.18 Red Alert for Rainfall 198
7.19 Overview Map of the Base Interface 208
7.20 Areas of Regio I 209
7.21 Insulae of Regio I 210
7.22 Global SA Per Scenario Barplot 223
7.23 Perception SA Per Scenario Barplot 223
7.24 Comprehension SA Per Scenario Barplot 224
7.25 Projection SA Per Scenario Barplot 225
7.26 Scenario 1 Score Barplot by Level 227
7.27 Scenario 2 Score Barplot by Level 228
7.28 Scenario 3 Score Barplot by Level 229
7.29 Scenario 4 Score Barplot by Level 229
7.30 Scenario 5 Score Barplot by Level 230

List of Tables

3.1 Summary of Approaches and Results on Movable Heritage . . 38
3.2 Summary of Approaches and Results on Immovable Heritage . 50

4.1 Susceptibility Analysis Criteria and Weights 72
4.2 Evaluation Scale for Susceptibility Subcriteria 73

5.1 IMU, RGB Camera and Platform Specifications 93
5.2 Test results from the Missing Completely Random Error Model 117
5.3 Test results from the Consecutive Error Model 118
5.4 Granulation Method Test Results from the Consecutive Error Model 119
5.5 Test Results against Neural Network-based Methods - High Correlation Version 119
5.6 Test Results against Neural Network-based Methods - Moderate Correlation Version 119

6.1 Division of Images and Labels into Train/Validation, and Test Sets. 142
6.2 Comparison between Test Set Results of YOLOV5-l and YOLOV8-l 147
6.3 F1-Score, Precision and Recall in Cross-Validation and Test on YOLOV5-l 148
6.4 PompeiiTile Dataset Results 160
6.5 PompeiiFlat Dataset Results 160
6.6 HydraML (Cos) Pompeii Dataset 167

7.1 JSON to OWL Conversion Table 190
7.2 Scenario 1 Questionnaire. 212
7.3 Scenario 2 Questionnaire. 213
7.4 Scenario 3 Questionnaire. 214
7.5 Scenario 4 Questionnaire. 215
7.6 Scenario 5 Questionnaire. 217

7.7 Global OLS-ANOVA p-values 224
7.8 Per-scenario ANOVA OLS VERGIL vs Baseline 231

Chapter 1

Introduction

"To be ignorant of what occurred before you were born is to remain always a child. For what is the worth of human life, unless it is woven into the life of our ancestors by the records of history?"

— Marco Tullio Cicerone

Cultural heritage sustains cultural identity, the economy, and the attractiveness of a region, serving as a primary vehicle for transmitting tangible, natural, and intangible assets to future generations [1]. Cultural tourism, rooted in the 18th century and consolidated in its modern form since the 1970s, centers on the exploration of natural landscapes and cultural heritage. Tourism linked to a country's cultural assets is a significant economic driver, especially where such heritage is abundant [2], [3].

In this thesis, we define *Cultural Heritage sites*, such as archaeological parks, museums, and historic structures, as *Critical Infrastructure* that should be preserved for both their social and economic value. Indeed, the European Union defines critical infrastructure as: *an asset, system or part thereof located in Member States which is essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact in a Member State as a result of the failure to maintain those functions* [4] [5]; while the United States of America defines critical infrastructure as: *systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters.* [6] [7]. Given these definitions, Cultural heritage sites can be considered critical infrastructure since they are essential to vital societal functions and to social and economic well-being. They sustain identity, education, and social cohesion, and they often support tourism

revenues and local employment, so their disruption or destruction can produce significant, long-lasting impacts. Because these sites are also irreplaceable, the consequences of damage are often irreversible, which aligns with both the EU and US ideas of infrastructure whose loss would seriously undermine security, prosperity, and public well-being.

The protection of such Cultural Heritage Critical Infrastructures (CHCIs) also acts as a strategic activity in order to satisfy the 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, which provides a shared blueprint for peace and prosperity for people and the planet, now and into the future. At its core are the 17 Sustainable Development Goals (SDGs), which are an urgent call for action by all countries, developed and developing alike, in a global partnership (`https://sdgs.un.org/goals`). They recognize that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and stimulate economic growth, all while tackling climate change and working to preserve our oceans and forests.

A serious plan for the preservation of cultural heritage is capable of supporting several of the 17 SDGs, in particular:

- **SDG 4 — Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.**
  Safeguarding CHCIs preserves history and knowledge for future generations, ensuring they can be studied directly, easily, and at high quality. If the past is not preserved, knowledge is lost, undermining culture within society itself.

- **SDG 8 — Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.**
  Well-managed CHCIs sustain local economies and quality jobs in conservation and tourism. Treating them as critical infrastructure prioritizes safe access, lifecycle maintenance, and visitor management that preserves value while preventing overuse. Actively supporting operators with new technologies makes their work easier, faster, and more effective. While some roles may become less necessary over time, new expert positions are emerging to complement existing ones, fostering both economic and social development.

- **SDG 9 — Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation.**
  Since cultural heritage is a domain involving critical infrastructures, they should be made resilient. Resilient CHCIs can be achieved by adopting emerging technologies and paradigms such as the Internet of Things

(IoT), artificial intelligence (AI), digital twins (DT), and structural health monitoring (SHM).

- **SDG 11 — Make cities and human settlements inclusive, safe, resilient and sustainable.**
  Safeguarding CHCIs sustains urban identity, the continuity of social life and the resilience of the territory. Investing in protection, through risk assessment, preventive conservation, and emergency planning, strengthens urban resilience and supports inclusive place-making. Technological development of CH sites, coupled with their preservation and sustainable economic use, protects the revenues of the institutions, territories, and cities that manage them, creating added value.

- **SDG 13 — Take urgent action to combat climate change and its impacts.**
  The adoption of new technologies and paradigms for protecting CHCIs can reduce the impact of climate change and adverse natural events, helping to prevent disasters. This is especially important in cultural heritage, where preservation also depends on how sites themselves treat the environment, so as not to worsen underlying risk conditions. For this reason, the enabling technologies used to support preservation must be sustainable and environmentally responsible.

- **SDG 17 — Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development.**
  Safeguarding CHCIs requires multi-stakeholder cooperation across government, academia, the private sector, and civil society, sustained by interoperable standards, shared data, and joint financing mechanisms for protection and recovery.

The efficient protection of CHCIs involves optimizing multiple steps and aspects, and leveraging a range of technologies. First, understanding the condition of CHCIs is fundamental [8]. Traditional monitoring procedures are often insufficient in cultural or natural heritage contexts, not least because a site can be difficult to access. Consequently, extensive data collection and multidisciplinary expertise are required [9], [10]. Today, these procedures can be improved through innovative paradigms and tools [11], which support the identification of risk factors, the design of risk-management strategies, and the tracking of maintenance activities to inform decision-making. IoT serves as a paradigm for collecting real-time data across different areas and aspects of CHCIs. A variety of sensors and devices are distributed throughout the physical asset to gather information that can be analysed subsequently.

To achieve this, systems may employ robots capable of performing a range of tasks and collaborating with humans to deliver more efficient, proactive monitoring and maintenance.

These robots, named *cobots*, are designed to coexist and work in the same environment as humans, prioritizing safety and efficiency rather than outright replacement; they typically handle repetitive, precision tasks so that humans can focus on problem-solving. In this way, humans and robots realize Human–Robot Collaboration (HRC), working together toward a common goal, such as identifying maintenance issues and, more broadly, safeguarding the CHCI. Multiple ground and aerial cobots can be deployed at CHCIs to capture high-resolution imagery, assess on-site conditions, and intervene when needed. These cobots can coordinate their inspections and supply valuable data to be integrated with readings from other sensors. They can also carry out repairs or direct interventions on-site, either autonomously or in concert with human operators [12] [13]. In the CH domain, such cobots can take the form of ground robots, either wheeled or quadrupedal, which are widely employed because they can access narrow and otherwise inaccessible areas. However, Unmanned Aerial Vehicles (UAVs), or more generally aerial drones, are used even more extensively. UAVs can perform surveys and capture high-resolution imagery that supports photogrammetric workflows to produce georeferenced orthophotos or 3D models of a site, potentially complemented by technologies such as LiDAR.

The outputs of such cobots are therefore complex virtual 2D and 3D objects, which help to characterize conditions within the CHCI and support operators' decision-making. However, the data collected during surveys must be analysed, and in large, heterogeneous environments this can be a lengthy and complex process if performed manually. AI can accelerate and enhance these analyses [14] [15]. AI supports CHCI operators in preservation and conservation tasks, enabling the early detection of issues and quick responses to unforeseen risks [15]. Predictive models and machine learning (ML) algorithms provide robust decision support in today's cultural landscape [16], [17]. By processing data from varied and heterogeneous sources, these tools detect trends and areas of interest, allowing managers and experts to adopt informed protective measures and tailor strategies to each context, which is invariably case-specific [18], [19].

To support the preservation of cultural heritage as critical infrastructure, our work focuses on implementing new paradigms that leverage recent technologies to strengthen protection and deliver tangible benefits.

Below, section 1.1 will outline the challenges and objectives of this work; section 1.2 will explain the research approach used; section 1.3 will declare the contributions of this work; section 1.4 will then explain how the thesis will be structured.

1.1 Challenges and Objective

This work also stems from a collaboration with the Pompeii archaeological site. In recent years, the site has undergone a major digital transformation, initiated primarily by the Smart@Pompeii project (`https://pompeiisites.org/archivio-progetti-e-ricerca/smart-pompei/`), through which it deployed sensor-equipped cameras and devices to monitor the site environment, as well as cobots for ground and aerial inspections. As part of this project, the site acquired a dog robot named *Spot* from *Boston Dynamics* to patrol the grounds and navigate Pompeii's uneven terrain, along with an aerial drone, a *DJI 300 RTK Matrix* equipped with a high-definition RGB camera, the *Zenmuse L1*. Even before our research began, operators were using this aerial cobot to survey the entire site and photograph specific areas. Approximately once a month, they conducted a full flyover to capture georeferenced images of the whole area, which were then merged into a site-wide orthophoto using photogrammetry software. Human operators subsequently inspected the orthophoto to identify maintenance issues, such as weedy vegetation, damaged conduits, and, above all, broken roof tiles on various structures. Because heavy rain or storms can drive water through broken roofs and damage interiors, artifacts, and frescoes, promptly resolving these issues is essential.

This hybrid, collaborative workflow, where drones collect data and humans manually verify the findings, introduces several challenges:

- Since orthophotos are generated on a monthly basis, operators remain blind until the next flight;
- The identification of maintenance issues is very slow, as operators must manually inspect the entire site, zone by zone;
- The identification of maintenance issues is error-prone, as operators may overlook issues due to distraction, fatigue, preconceptions, or similar factors;
- Such identification requires resources and human effort that are not always available or that are too expensive.

The identification of maintenance issues can be facilitated and accelerated by automating their recognition using AI techniques, thereby reducing human workload. However, in this domain, identifying such issues is itself a complex task. Both human and artificial agents may struggle to distinguish different areas of an archaeological park in photos, since these heterogeneous contexts combine natural landscapes with urban ruins, and it is not always clear which zones actually require maintenance. At a glance, for example, a damaged roof

might belong to an ancient structure that is intentionally in ruins, or it could be part of a modern reconstruction in Roman style that must be repaired to protect artifacts and frescoes inside. Similarly, vegetation overgrowth may be a serious issue in dense, built-up areas, while being entirely normal in other parts of the park where structures are sparse.

Considerable work is therefore needed to develop ad hoc models that account not only for accuracy, but also for energy and resource consumption, in order to minimize environmental impact and ensure that monitoring activities do not inadvertently compromise the preservation of the assets themselves.

In this context, we found that full automation is not feasible, since an incorrect decision could have disastrous consequences. Because we are dealing with critical infrastructure, final decisions on what actions to take and what to maintain must remain with human operators. Any intervention must be justified, and priorities within maintenance plans must be assessed by domain experts. We therefore operate in a setting where close interaction among the physical asset to be preserved, the technologies used to monitor and manage it, and the people who work with it is essential. For this reason, we propose to consider a general CHCI as a Cyber-Physical-Social System (CPSS), i.e., a system that integrates digital technologies with physical entities, sensors and actuators, and human interaction with both machines and other humans. The value of the CPSS paradigm lies in its well-established foundation in the literature, which enables the interaction among all actors in the system to be formalized and implemented more effectively. Although CPSS have been adopted in other domains, we argue that this perspective is also useful and applicable to cultural heritage, as we will aim to demonstrate in this thesis.

At the same time, to make sound decisions, users must always remain aware of what is happening. In practice, their Situation Awareness (SA) must be supported. SA is defined as: *the ability of a human or artificial agent to understand what is happening in the current environment to support informed, adaptive, and resilient decisions* [20]. In fact, several studies have shown that, under the same conditions, a higher SA among human operators leads them to make better decisions on average [20] [21] [22]. Especially in cases like this, where the user cannot be replaced but some operations are automated, it is crucial that the user stays in the loop of system actions, so they do not lose awareness and are always ready to intervene. Users may also struggle to track everything that is occurring; a supportive system can therefore be essential. When operators must attend to many aspects at once, dividing attention is difficult. If they are focused on a specific goal, they tend to filter out and ignore elements that are not directly related to that goal, but can be important for the safeguard of the CHCI. Thus, the system must draw the user's attention when necessary.

At the same time, maintaining a high level of situation awareness while op-

erators are waiting for the next survey can be achieved by exploiting contextual information such as weather conditions and visitor flows. This information helps keep vulnerability assessments of different site areas up to date and supports understanding of how these conditions are evolving based on the most recent survey data. This aspect is particularly important in contexts such as the archaeological site of Pompeii, where operators are effectively blind to new maintenance issues for an entire month until the next aerial orthophoto becomes available.

Following these considerations, the objective of this thesis is to develop a methodological approach that enables users to understand how, where, and when to intervene to protect CHCIs.

To pursue this overall objective, this thesis addresses the following specific research challenges in safeguarding CH:

Challenge 1. It is not possible to automate everything; a human operator always makes the final decision on maintenance actions. That operator must remain constantly aware of what is happening around them;

Challenge 2. Human operators must be able to recognize risk situations in order to implement effective maintenance plans;

Challenge 3. Human operators may encounter low-quality or missing sensor data. Because data reliability strongly influences SA, safeguards must be put in place to support the perception phase;

Challenge 4. Perceiving and understanding maintenance issues in a CHCI is a slow, complex task for human operators. Problems must be detected before they cause serious damage to the infrastructure;

Challenge 5. Human operators handling maintenance issues face *data overload* and *attentional tunneling*, which can compromise their SA. Under these conditions, they may miss the most important environmental signals and fail to understand what is happening.

To address these research challenges and achieve the overall goal, this thesis defines the following research objectives, along with the activities and tasks required to fulfill them.

Objective 1. Analyse state-of-the-art approaches and techniques to support human operators' decision-making in the maintenance of CHCIs:

Task 1.1. Understand current approaches, techniques, and regulations for maintenance activities in CHCIs;

Task 1.2. Understand how SA demons affect SA assessment processes in CH, and identify the key factors that positively or negatively influence them;

Task 1.3. Analyse the key research works on SA applied to CH.

Objective 2. Define situations within a CHCI and manage them through a novel methodological approach that supports users' decision-making:

Task 2.1. Analysis of the current approaches and indices used to evaluate the vulnerability of CHCIs;

Task 2.2. Define a novel vulnerability index combining on-site maintenance issues with environmental contextual data;

Task 2.3. Define a novel Situation-aware Cyber-Physical-Social System methodological approach for cultural heritage protection that formalizes interactions among the physical CHCI, the technologies used for its preservation, and the operators and visitors who interact with it, thereby supporting maintenance-related decision-making. This SA-based approach enables users to remain informed about ongoing conditions even during the extended intervals between on-site surveys. By maintaining and sharing awareness of what is happening, the CPSS allows users to concentrate on the reasoning processes required to make well-informed decisions;

Task 2.4. Conduct a proof of concept to assess the feasibility of the approach in real-world CHCI settings.

Objective 3. Address the challenge of low-quality and missing sensor data by reconstructing missing information and reducing perception-level errors:

Task 3.1. Analysis of the state-of-the-art of data imputation approaches;

Task 3.2. Define a novel lightweight data imputation approach for small physical and virtual sensor networks, typical of archaeological sites and museums;

Objective 4. Automate the identification of maintenance issues in CHCIs using a sustainable, energy-efficient AI approach:

Task 4.1. Analysis of AI approaches for CH protection;

Task 4.2. Define a new sustainable classification approach to identify maintenance issues on CHCI roofs.

Objective 5. Ensuring that operators are consistently prepared to respond to emergencies, while maximizing their performance both in routine activities and when facing imminent risks:

Task 5.1. Design and develop a user-oriented platform to support human operators in maintaining archaeological sites;

Task 5.2. Evaluate the approach in a real-world archaeological site application.

1.2 Approach

The objectives outlined in the previous section are pursued through the research approach illustrated in Figure 1.1.

The first step was to analyse current approaches and regulations used for CHCI maintenance. This knowledge was then used to identify the steps commonly performed today, resulting in a practical lifecycle for CHCI maintenance. Next, we investigated the SA demons that negatively affect situation understanding in these contexts, particularly data overload and attentional tunneling. We also reviewed the state-of-the-art on SA in cultural heritage, with a specific focus on CHCI maintenance, and finally we examined current approaches for evaluating the vulnerability of structures and areas within CHCIs.

We formally define situations in CHCIs using a novel vulnerability index for cultural heritage maintenance. This index identifies risk situations across site areas by combining detected maintenance issues with contextual information.

To support human operators in their tasks, we must manage complex systems that encompass CHCIs, technologies for handling risk situations, and people interacting with the physical world. For this reason, we propose a general methodological Situation-aware Cyber-Physical-Social System (SA-CPSS) approach for Cultural Heritage Protection, designed to help operators manage these risk situations and applicable across a range of cultural heritage contexts. This situation-aware approach is enabled by integrating a Cyber-Layer that automates on-site surveys and maintenance-issue recognition as much as possible, keeping operators informed about ongoing conditions even during the extended intervals between on-site surveys that are typical of the cultural heritage domain. The CPSS can then share its situational understanding with

the operator, reducing workload and allowing them to focus on critical decisions that cannot be fully automated, such as which actions to take and when to implement them.

We conducted a feasibility study to assess the validity of the approach by applying it to three case studies involving the archaeological sites of Pompeii, Paestum, and the Colosseum. The results were encouraging and motivated us to implement the model and further refine it.

Although this thesis evaluates an archaeological-site scenario, driven by our collaboration with the Pompeii site, the proposed methodological CPSS approach can be transferred to other cultural heritage protection domains. Any interested stakeholder could create similar experiences and products, as long as the modules described are defined and the interactions among them are specified in detail.

To ensure that the information retrieved from CHCI surveys is reliable and usable, we proposed a novel lightweight data-imputation technique to reconstruct missing values in small physical and virtual sensor networks within the sensing module responsible for CHCI data acquisition and preprocessing. After analysing traditional and recent imputation methods based on machine learning and deep learning, we proposed a granular association-rule imputation method that combines rule-based techniques with machine learning to impute data in small networks of physical or virtual sensors. In this way, we obtained a CPSS that, when needed, can reconstruct missing data in IoT and edge-computing settings, where computational and storage resources are often insufficient for deep learning, while achieving comparable performance and improving upon traditional machine-learning approaches.

The situation-aware Cyber-Layer must also enable the automatic identification of maintenance issues in the CHCI. In the case of archaeological sites, as indicated by experts, the most pressing maintenance issues concern roof conditions. Roofs are essential for the structural stability of buildings and, above all, for protecting the artefacts and frescoes inside.

To address the challenges described in Section 1.1, including the difficulty of classifying roofs in archaeological sites and the need to reduce energy consumption and improve sustainability, we defined a novel hybrid approach that uses AutoML to select deep-learning models for feature extraction and machine-learning methods for damage classification on roof data from the Pompeii archaeological site. This approach aims to achieve performance comparable to, or better than, end-to-end deep-learning solutions while requiring shorter training times and lower energy consumption. In addition, compression techniques applied during image acquisition and feature generation reduce dimensionality, training times, and noise, helping to optimise both efficiency and performance.

Active operator support is enabled by a user-centered interface that makes

the CPSS features available to the operator. A methodological approach with an adaptive interface can, regardless of the user's current goal, keep them informed about imminent emergencies by considering not only the identified maintenance issues but also contextual data about the environment and visitors, in order to advise them on the most appropriate actions to take.

As part of this thesis, we applied the SA-CPSS methodological approach to the Pompeii Archaeological Park by developing the VERGIL platform, which supports park operators in decision-making when creating maintenance plans.

To assess gains in operators' situation awareness across different scenarios, we used the Situation Awareness Global Assessment Technique (SAGAT) [23] to evaluate human operators' performance in dedicated use scenarios at the archaeological site of Pompeii. The results were then analysed and discussed.

1.3 Contributions

The main contributions of our work are as follows:

Contribution 1. An overview of current maintenance procedures and vulnerability assessment techniques in CH, resulting in the proposed CHCI Maintenance Activity Lifecycle;

Contribution 2. An overview of state-of-the-art approaches, techniques, and technologies in AI applied to the safeguarding of CH;

Contribution 3. A novel methodological SA-CPSS approach for Cultural Heritage Protection;

Contribution 4. A novel goal-driven Cultural Heritage Vulnerability Index (CHVI) to define situations within CHCIs;

Contribution 5. A novel lightweight data-imputation approach for small physical and virtual sensor networks;

Contribution 6. A new sustainable classification approach to identify maintenance issues on CHCI roofs;

Contribution 7. A novel SA-oriented platform for archaeological site protection. The platform is an instance of the SA-CPSS deployed at the archaeological site of Pompeii and features an adaptive interface based on a novel Oscillatory Goal Suggestion approach. It supports human operators in both routine and emergency maintenance activities through this adaptive interface;

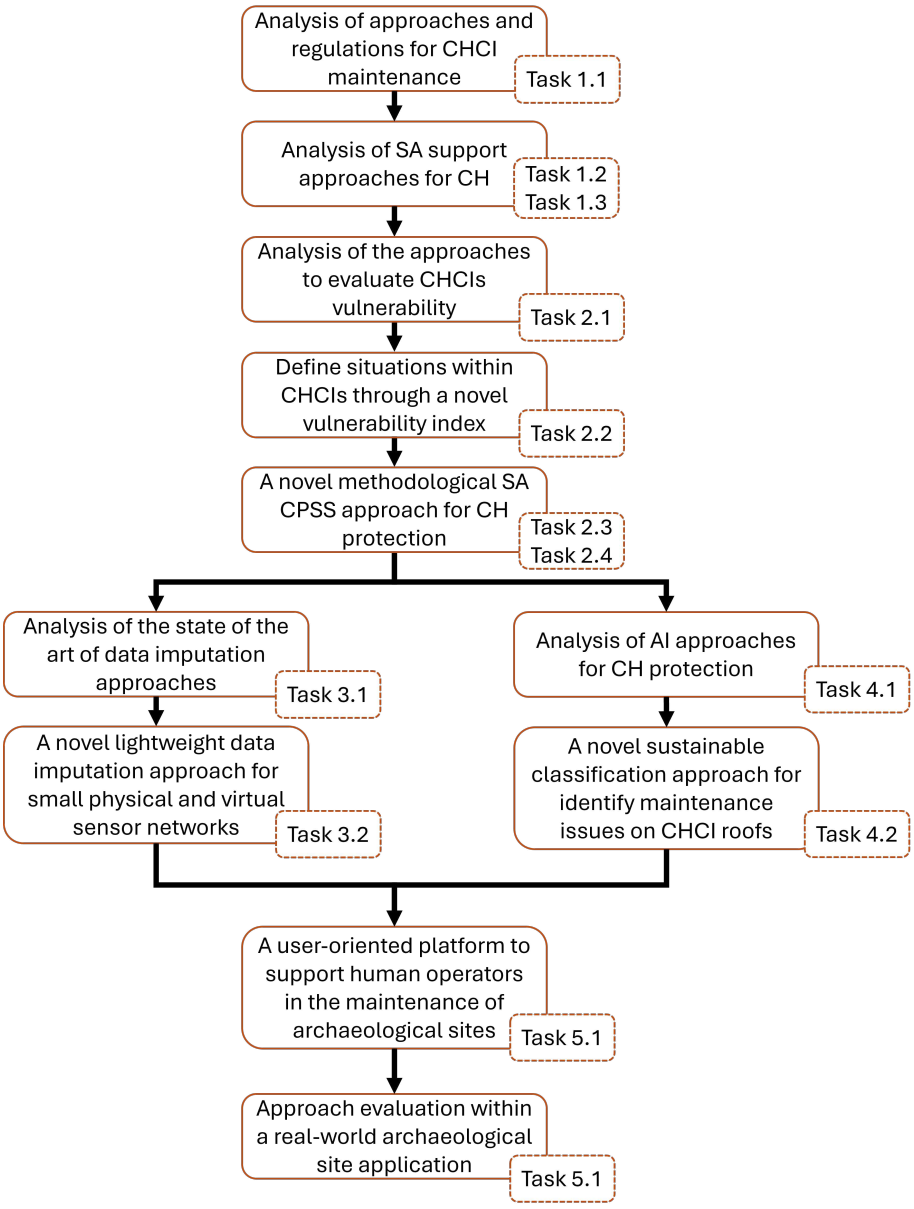


Figure 1.1: Research Approach

The previously stated contributions were evaluated through the following procedures:

- A numerical evaluation of our data-imputation approach using real IoT datasets with heterogeneous characteristics in terms of size and inter-sensor correlation;
- A numerical evaluation of our classification approach for identifying maintenance issues on CHCI roofs using real-world datasets from the archaeological site of Pompeii;
- A SAGAT-based evaluation of the novel methodological SA-CPSS in real scenarios at the Pompeii archaeological site. This evaluation assessed human users' performance while they carried out dedicated usage scenarios.

1.4 Thesis Outline

Chapter 2 introduces the theoretical foundations of predictive maintenance, CH maintenance practices, Cyber-Physical-Social Systems, digital twins, and situation awareness, with particular attention to SA-related demons and errors.

Chapter 3 provides a comprehensive overview of the state-of-the-art in AI applied to CH. The analysis identifies the main techniques adopted for CH protection and introduces a taxonomy covering the types of heritage to be safeguarded, the data sources considered, and the issues addressed, with particular emphasis on the detection of maintenance issues across different sites and artifacts.

Chapter 4 proposes a novel SA-CPSS methodological approach for Cultural Heritage Protection designed to improve the decision-making capabilities of human operators responsible for CHCI maintenance. We define a novel goal-driven Cultural Heritage Vulnerability Index (CHVI) to identify and formalise situations within CHCIs. The chapter includes a proof of concept to assess the feasibility of the approach and to illustrate how it works. This proof of concept is based on case studies at Pompeii, Paestum, and the Colosseum, three sites with distinct characteristics.

Following the CPSS architecture, the next chapters focus on its main components.

Chapter 5 describes the system's physical and virtual sensors, the adopted data sources, and the preprocessing pipeline. It covers both human-driven procedures, such as producing site orthophotos, and automated processes, such as data cleaning. The chapter also introduces a novel lightweight data-imputation approach for small physical and virtual sensor networks, named

GARDA: a Granular Association-rule-based Data imputation Approach, and evaluates it on several IoT datasets.

Chapter 6 addresses the automation of maintenance-issue identification in CHCIs using a sustainable, energy-efficient AI approach. Two alternative solutions are presented. The first is *FAUNO*, a YOLO-based system for detecting maintenance issues at archaeological sites. We then focus on assessing roof conditions through *HydraML: Hybrid Damaged Roofs AutoML Classifier for Cultural Heritage*, a sustainability-oriented AutoML method that operates directly on photographs of known structures. This approach assumes that roof locations are already known, since the mapping of archaeological sites typically changes very slowly over time.

Chapter 7 proposes *VERGIL*, a novel SA-oriented platform for archaeological site protection. The platform includes an adaptive interface with a novel Oscillatory Goal Suggestion approach that supports human operators in both routine and emergency maintenance activities. The interface can progressively alert users when an emergency requires their attention. VERGIL integrates the contributions described above across its modules and applies them to a real-world setting at the archaeological site of Pompeii. The chapter concludes with an overall assessment of the proposed methodological approach using SAGAT. This evaluation assessed human users' situation awareness and decision-making performance while they carried out dedicated usage scenarios related to the archaeological site of Pompeii.

Chapter 8 concludes the thesis by summarising the main results and outlining future developments and research directions.

Chapter 2

Theoretical Background

"In the fields of observation, chance favours only the prepared mind."

— Louis Pasteur, *Lecture, University of Lille, 1854*

In this chapter, we present the conceptual and technological background necessary to understand our proposal and to situate it within the broader research landscape. We first introduce the fundamental concepts and definitions of predictive maintenance, with a specific focus on its application in the cultural heritage domain. In particular, we discuss how predictive maintenance differs from traditional corrective and preventive approaches and which technologies have emerged to enable this new type of maintenance through data acquisition and analysis.

We provide an in-depth overview of maintenance practices in CH, reviewing current standards, guidelines, and real-world procedures. On this basis, we propose a summary of these practices, introducing our reference lifecycle tailored to CH entities. This lifecycle highlights the main phases, actors, information flows, and decision points that characterise maintenance activities, and serves as a reference framework for the subsequent design of our approach.

We outline the main concepts behind Digital Twins and CPSS for critical infrastructures, including CH entities considered as complex systems embedded in their physical and social context. We discuss enabling technologies, data acquisition and integration challenges, and the role of these paradigms in supporting monitoring, prediction, and decision-making processes for the preservation of CH.

We introduce the core ideas of SA as a key concept for human and CPSS interaction in complex monitoring and control environments. We cover the Endsley model of SA, cognitive and human factors influencing SA, system design considerations to support and maintain appropriate levels of SA, and

common SA *demons* and errors that can undermine operators' understanding and decision-making. These elements are essential to understand both the novelty and the design choices of our proposal, especially regarding how information is presented, interpreted, and acted upon by different stakeholders involved in CH maintenance.

Parts of this chapter have been previously published in:

- Francesco Colace, Giuseppe D'Aniello, Massimo De Santo, Rosario Gaeta, Gabriel Zuchtriegel. *Situation-aware Cyber–Physical–Social System for Cultural Heritage*, Intelligent Systems with Applications. 2025. Volume 27. `https://doi.org/10.1016/j.iswa.2025.200544`.
- Emanuele Bellini, Giuseppe D'Aniello, Francesco Flammini, Rosario Gaeta. *Situation Awareness for Cyber Resilience: A review*. International Journal of Critical Infrastructure Protection. 2025. Volume 49. `https://doi.org/10.1016/j.ijcip.2025.100755`.

2.1 Predictive Maintenance

Maintenance activities play an important role in various sectors, especially in industry, since their cost may represent a significant part of the production costs. A good maintenance plan allows one to avoid unexpected production stops, to reduce the costs, and to increase the lifetime of industrial machines. However, maintenance is an essential activity in all sectors in which safeguarding certain entities must be carried out, such as CH: if entities are not continuously monitored and preserved, irreparable damage may occur [24].

In such a context, maintenance activities have evolved greatly over time. The most primitive strategy is the corrective maintenance, also known as *run-to-failure*, which consists of replacing or repairing a part only when it is damaged. Due to the high costs of having unexpected production stops, CM evolved into proactive approaches. Such proactive approaches are also fundamental in the CH domain since it may not always be possible to properly repair a structure or artifact and restore it to its original condition if the damage is too great. However, the first approach to emerge was the preventive maintenance, which involves periodic inspections by experts of the field. This replacement is made at equally spaced intervals, generally provided by the experts. This may lead to an unnecessary restoration at best, or a delayed intervention at worst, causing

more damage [25]. However, nowadays the inspections made by experts can be automatized or semi-automatized with the help of new technologies and paradigms like IoT, UAVs, AI, novel sensors and devices capable of measuring, monitoring, and processing signals that represent physical parameters of industrial equipment, such as acoustic signals, current, voltage, temperatures, forces, vibrations, and others. Predictive maintenance uses these tools to automatically collect data on a problem, analyse it and predict possible future evolutions of the phenomenon, so as to understand when it is necessary and best to carry out maintenance [26].

2.1.1 Maintenance in Cultural Heritage

We believe it is essential to highlight maintenance procedures in the cultural heritage domain and the methodologies used to assess vulnerability situations in this context. We therefore summarise the maintenance procedures in our reference maintenance lifecycle and introduce recent approaches to measuring vulnerability in the field of cultural heritage.

2.1.1.1 Cultural Heritage Maintenance Lifecycle

In the CH domain maintenance activities are fundamental for the CH field, and, for this reason, international institutions and organizations have established a set of regulations and procedures for performing such activities. Any system in this context that wants to support maintenance activities related to CH entities has to deal with such procedures, which is why it is useful to summarise them as the proposed CH Asset Maintenance Activity Lifecycle in Figure 2.1, defined from the integration of various vulnerability assessment frameworks [27] [28] [29] and from European Projects [30], focusing with particular attention to the EU and Italian experiences [31] [32] [33]. The lifecycle is composed of the following six steps:

- **S1 Preliminary Survey:** the preliminary survey refers to the procedures that must be generally performed when there is the discovery of an entity:
  - Preliminary Data Collection: gather of historical, bibliographic and cartographic information about the site;
  - Field Reconnaissance: visit of the site to understand and to report current conditions;
  - Geomorphological Analysis: geomorphological study to understand the context (where possible);

  - Photo Interpretation: identification of structures or anomalies (where possible) starting from aerial photographs;
  - Evaluation of Constraints: check limitation regarding environmental, historical, landscape or material constraints.

- **S2 Detailed Analysis:** activities carried out to analyse the state of the CH asset:
  - Archaeological Excavations: excavations activities to find ancient structures and artifacts (where possible);
  - Documentation: retrieval through photographs, drawings and descriptions of the findings;
  - Extensive Investigations: empirical and instrumental checks, with detailed analysis of the relief to find cracks, deterioration and problems, in order to document it;
  - Dating: use of techniques such as radiocarbon dating to determine the age of the finds;
  - Materials Analysis: study of materials to determine composition and state of preservation.

- **S3 Risk Assessment:** activities to assess the risk and vulnerability of the structures and art pieces of the CH asset:
  - Structural Analysis: evaluation of the stability of existing structures (where possible);
  - Environmental Monitoring: monitoring of environmental factors such as humidity, temperature, pollution, etc. We can also consider as ambient monitoring the anthropic flux monitoring;
  - Seismic Risk Assessment: analysis of site vulnerabilities to earthquakes and climate change (where possible).

- **S4 Intervention Planning:** once analysis and risk assessment have been performed, the intervention plan must be drawn up:
  - Prioritization of Interventions: identification of parts and areas in need of urgent interventions.
  - Design of Interventions: preparation of detailed plans for restoration and conservation.
  - Resource Allocation: planning for the necessary financial and human resources.

- **S5 Intervention execution:** the intervention plan is put in place:

  - Restoration: restoration work to repair and preserve structures and artifacts.
  - Preventive Conservation: measures to prevent further damage, such as moisture control and protection from the environmental elements.
  - Routine Maintenance: regular maintenance activities to keep the asset in good condition, e.g. removal of dust deposits, dirt, mosses and lichens from surfaces, drainage channels, etc, restoration of minor injuries, replacement of damaged or missing elements, removal of unsafe parts and installation of temporary protection systems, checking the operating status of windows and doors, fixtures, safety devices.

- **S6 Monitoring Survey:** at this point, the survey phase begins again:
  - Regular Inspections: periodic inspections to monitor the state of conservation;
  - Documentation Update: continuous recording of the asset conditions and interventions carried out;
  - Evaluation of Effectiveness: analysis of the results of interventions to improve future conservation strategies.

Figure 2.1: Cultural Heritage Maintenance Lifecycle

2.1.1.2 Cultural Heritage Vulnerability Index

The maintenance activity planning is generally recognized as the most complex one. For maintenance planning to be performed properly, practitioners must rely on rigorous metrics for vulnerability assessment in CH. However, a standard for the vulnerability assessment indices has not yet been identified [34] [35] [36], but several definitions have been proposed [37] [38]. Recently Ravan et al. [27] proposed a Cultural Heritage Vulnerability Index (CHVI), defined as:

$$CHVI_r = \text{Susceptibility}^{1/2} \times \text{Lack of Coping \& Adaptive Capacity}^{1/2} \tag{2.1}$$

where the terms are:

- **Susceptibility:** defined on the basis of a heritage site's physical characteristics, reflecting the performance of its structural and material features against the effects of natural hazards;

- **Coping:** is *the combination of all the strengths, attributes and resources available within an organization, community or society to manage and reduce disaster risks and strengthen resilience* [39];

- **Adaptive Capacity:** is *the ability of systems, institutions, humans and other organisms to adjust to potential damage, to take advantage of opportunities, or to respond to consequences* [40].

In a subsequent work, Cacciotti et al. [34], proposed a methodology for evaluating the vulnerability of built heritage and cultural landscape exposed to hydrometeorological hazards, defining a vulnerability index as:

$$CHVI_c = Exposure + Susceptibility - Resilience \tag{2.2}$$

where exposure represents the conditions *of people, infrastructure, housing, production capacities and other tangible human assets located in hazard-prone areas* [41], susceptibility is the same defined by [27], and resilience identifies *the ability of a system exposed to hazards to resist, absorb, accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner* [41]. Such metrics are associated with a series of criteria with weights and possible values, contributing to the final CHVI.

More detailed explanations of how to calculate this index and its applications are provided in Chapter 4.

2.2 Cyber-Physical-Social Systems and Digital-Twins

From the previous discussion, we understand that predictive maintenance is fundamental in the context of critical infrastructure protection, as a CH entity. Enabling approaches for safeguarding such types of infrastructures are Cyber-Physical Systems and Digital Twins.

In particular, a Cyber-Physical System is a concept that integrates computational and physical processes. Cyber-Physical Systems are, in general, networked systems in which the computational (cyber) part is tightly integrated with the physical component [42]. When the integration is also extended to the social component, we can talk about Cyber-Physical-Social System (CPSS). CPSS is a combination of secure digital technologies with physical systems, sensors and actuators, and it additionally incorporates social aspects, represented, for example, by human interaction to facilitate automated and secure services to end-users and organisations [43].

A digital twin is defined as a virtual representation of a physical asset enabled through data and simulators for real-time prediction, optimization, monitoring, controlling, and improved decision-making [44] [45]. Digital Twins are essentially virtual copies of products, processes or services which encompass all the above qualities [46] [47]. To do so, the digital twins are equipped with data acquisition modules, with which they can map the physical object in real-time, reflecting every element and feature of the physical in the digital [46]. In this regard, the Digital Twin typically integrates IoT paradigms and technologies with BIM and HBIM, so that the real model is a replica not only visually but also of what happens in the object [48]. In this way, weather and environmental sensors, but also lidar, photogrammetry, laser scanning cameras and UAVs contribute to the creation of the digital replica [49]. However, the creation of a digital object mirroring the real thing is not sufficient for the identification of a digital twin as conceived, precisely because the digital twin does not only have a monitoring function. Instead, such a digital entity has the possibility of analysing the data arriving in real-time, so as to signal possible problems [50]. AI techniques can be used to detect problems but also to understand situations that could be problematic before they happen, and thus carry out correct and timely maintenance. The digital twin also has a simulation component, which allows the user to freely explore what might happen in the future, creating what-if scenarios and supporting the decisions of a human operator, whether in routine, urgent or planned maintenance [51].

Ultimately, Cyber-Physical-Social Systems and Digital Twins are technologies that are widely embedded in today's world, because any information system that controls processes or physical entities and is also closely related to interaction with humans, who play a role in control or use, can be seen as a

Cyber-Physical-Social system. Such a system can be viewed as the union of a physical twin and a digital twin, where the digital twin provides the enabling applications and services for monitoring, analysis, and simulation of scenarios within this complex system. In this context, digital twin elements can provide valuable support for the innovation and implementation of cyber-physical-social systems.

2.3 Situation Awareness

Situation Awareness (SA) is a concept related to humans that is influenced by their cognitive mechanisms and decision-making processes, as well as by the way they process information and by human factors. In common usage, SA means to understand what is happening around us at a specific moment to make a good decision with respect to our goal. Indeed, we can distinguish between SA intended as a state of knowledge of what is happening at a given moment, and the process of gaining and maintaining the SA by elaborating and understanding new information. To avoid ambiguities, we refer to the process of gaining SA as situation assessment. SA was first applied in military aviation [23] [52]. Over the years, however, SA support has expanded to any domain in which human operators must make decisions [53] [54] [55] [56].

A general definition of SA is given by Endsley: *Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.* [20] This definition explicitly encapsulates the three stages of SA:

- **Level 1 (L1) Perception:** the first level of SA is the perception, which refers to the identification of the status of the elements in the environment. Obviously, the elements to perceive depend on the domain of interest but, more importantly, on the goal of the users. Generally, perception of the elements is performed with various senses and considering various sources. The perception is often the most critical activity, indeed Jones and Endsley found that most of the SA errors occur at that level. Indeed, to detect all the needed data and information can be a huge issue that may be due to a shortage of data to record, so that it becomes difficult to derive the desired information, or to an overload of such data, which makes it difficult to grasp only the information that is really useful for our goal, eliminating information that is of little importance at a given time.

- **Level 2 (L2) Comprehension:** the second level of SA consists of un-

derstanding what the data coming from the perception mean in relation to goals. Such comprehension occurs by aggregating such data around the goal, but it is a complex activity because it often requires substantial domain expertise related to the domain, expertise that is necessary to build good mental models to interpret the situations.

- **Level 3 (L3) Projection:** the third level consists of understanding the possible evolution of the perceived elements with respect to the goal and the environment. A good L3 of SA must be based on a solid level of comprehension and on a good knowledge of how the system and the environment evolve. Even if the L1 is the most critical activity because it is most prone to error, projection is probably the most complex, as it requires a good understanding of the domain, of the situation and a great ability to project the status of many elements in the future. For this reason, experience related to the domain of interest plays a major role, because it enables the understanding of how the situation could evolve in the future and of how to be proactive in such a context.

The SA, made by the three levels considered, should be seen as a process that is preliminary to decision-making. In such a context, the three phases of the model should not be interpreted as three separate and sequential steps, but each level depends on and influences all the others, updating their knowledge iteratively [57]. The Endsley model for SA illustrated in Figure 2.2 considers SA itself as a necessary prerequisite for decision-making, which in turn is required to perform actions. In such a context, an action performed will produce an effect on the environment, modifying the situation. The Endsley model also highlights how these three aspects, situational awareness, decision-making and action execution are strictly connected with the aspects of the system used by an operator, such as system and interface design, workload, complexity and automation. In turn, aspects related to the person and the individual, such as goals and expectations, influence SA and decision-making, as do long-term and working memory, abilities, skills, experience, and training, which also influence action execution. From the Endsley model just analysed, it is possible to deduce precisely one of the main differences between the concept of SA and that of context awareness, which is widely confused in the literature. Context awareness is in fact a concept that is linked to a predominantly environmental awareness aspect, which does not take into account the cognitive and system design aspects that are instead fundamental in SA, because they influence every aspect of SA, decision-making and action execution. SA is a more evolved and complex concept than context awareness, which, however, makes extensive use of the context. In the following subsections, we will investigate in detail such cognitive and systemic aspects.

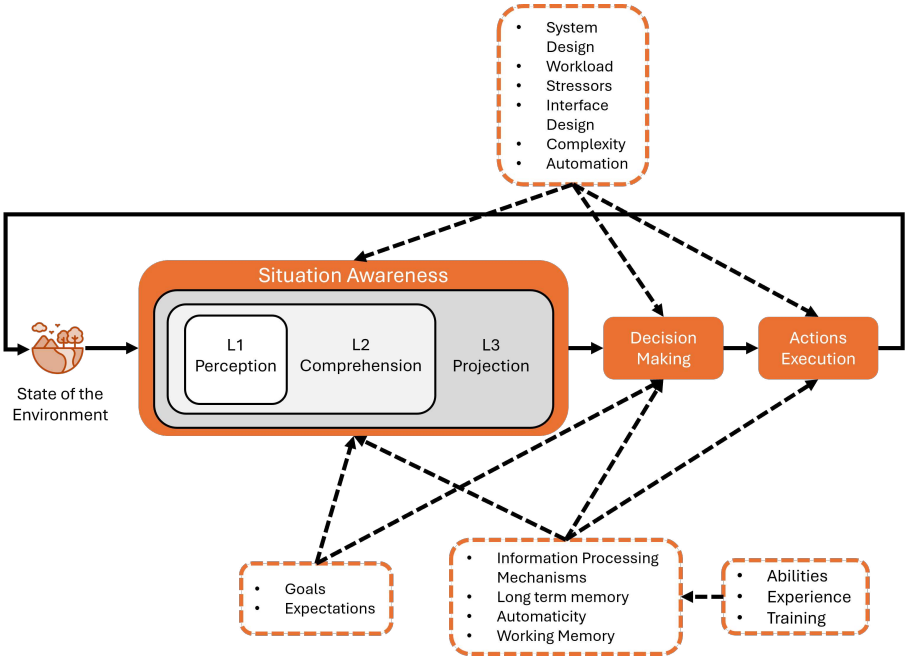


Figure 2.2: Endsley Model for Situation Awareness (adapted from [20])

2.3.1 Cognitive and Human Factors

As summarised by the Endsley model, the SA is influenced by various cognitive processes [20, 57–59]. In particular, the Working Memory (WM) stores information for short periods and it is fundamental for any type of activity performed by a user. These processes are affected by WM's limited capacity, in particular it can contain 7±2 pieces of information. Unfortunately, stored information tends to be lost quickly in the absence of focused concentration on maintaining it. This limitation can cause SA problems, as some important information that has been received can be forgotten and lost quickly. The Long Term Memory (LTM) is responsible for storing information for the long term. Mental models and schemata compose such Long Term Memory. Mental models are how we model the systems around us, how they will evolve, how they function and they are so important because they help people understand what information is important to consider and to predict what can happen next. If the mental models are not good, these mechanisms will not work. A schema represents the system in a certain situation, in a way that elements perceived from the environment are matched with the available schemata in a short time. In such a way, a person is able to classify and understand a well-known situation that is recurring. The script is the last structure of LTM, which can be defined as a sequence of

actions to perform in a given situation, so that, when a schema is identified, the script automatically triggers the sequence of actions to perform, speeding up the choice of a countermeasure. In addition, the user's goal is the main driver of SA assessment and decision-making. The SA assessment process is indeed goal-driven, since a goal identifies the information of the environment to which to pay attention. Indeed, the active goal pursued by an operator identifies what mental model has to be used. This makes it important for the user to focus on the correct goal since if such goal is wrong, fundamental information may not be correctly interpreted. In such a context, it is important to consider that individuals do not always use only information coming from the environment, but it is intrinsic to human behavior to have some expectations of what could be the state of elements and what will happen next, even if the user has not yet perceived anything in the short time. When having expectations, people do not need to acquire and process other information from the environment to project what will happen next and make a decision. Unfortunately, false expectations can lead people to misinterpretations of data and to miss some information. Lastly, automaticity allows people to react to what is perceived without using mental models or other cognitive processes, thus freeing up many mental resources for other tasks. This is a mechanism connected to instincts, innate or acquired as muscle memory. Unfortunately, in some situations, automaticity can be dangerous for the SA, as the information that is outside the loop of the automaticity may not be perceived.

2.3.2 Systems Design Factors

SA is also influenced by tasks and system features [20] since the design of a system and its interface strongly influences SA. Indeed, in general, not all the information acquired can be displayed on an interface since sensors could have some issues over time but also because showing all acquired information may not be the best option in many cases: human cognitive mechanisms are limited and data overload can be counterproductive. For this reason, the design of systems and interfaces should be user-centered to reduce the mental workload of an operator. Jones and Endsley in [59] provide some principles for designing systems and interfaces. In such a context, a key factor influencing SA is the stress. The problem is that typically systems aggravate the situation by introducing significant stressors, like physical (noise, vibration, heat, etc.) or social (fatigue and uncertainty) ones. Although a certain amount of stress can be good for the user's attention, too high a level of stress becomes counterproductive. Even if the workload exceeds what a human can process, SA may suffer. This excessive workload is often caused or made worse by the very systems, which

can be too complex. A proper design of such a system and training of the operators can foster the creation of good mental models. If, on the other hand, the system merely focuses on technological possibilities, it risks overloading the user with data and damaging their SA. Furthermore, the level of automation of a system can help in reducing the workload, but when too much automation is available, the out-of-the-loop syndrome may happen, causing a loss in the SA.

2.3.3 SA Demons

SA assessment can be generally made more problematic by the combined effect of cognitive mechanisms and system factors. A number of problems can arise from these factors, namely the *SA demons* [59], which should be considered when designing a SA system. In the following, each SA demon is reported as described by Endsley [59]:

- **Attentional tunneling:** In order to achieve a good level of SA, a human operator must often be able to pay attention to different tasks and information sources. In this regard, one of the demons is named attentional tunneling, which happens when the operator is able to pay attention to only a subset of the available information. In such a way, the operator does not have the overall picture, and the SA can be severely compromised;
- **Data overload:** In complex domains, if the amount of data that the operator must process and the refresh rate of such data is high, the operator has difficulty processing such data. This issue could be reduced through a good system design, understanding the needed data, and adopting suitable data analysis techniques;
- **Complexity:** A too-complex system with over-articulated features and rules may lead an operator to fail in constructing correct mental models. Training, user-experience and a good interface design can be fundamental to limit such issues;
- **Memory trap:** Humans have a limited space available in the working memory, so that certain information, even important information, is volatile, and can even create false memories in some cases, which can negatively affect SA. The use of correct mental models, as well as the experience, may help in freeing up the working memory;
- **Workload and stressors:** Excessive workload, stress, fatigue, and anxiety undermine the SA seriously. Such stressors can come from the

environment but can also be caused by the tasks or personal factors. Such stressors can also reduce the efficiency of the working memory, increasing the effect of the memory trap demon;

- **Wrong mental models:** If the operator applies an incorrect mental model, they may misinterpret the meaning of the perceived information and how the situation might evolve, leading to poor decisions. This often happens when a user is using a system for the first few times and, being accustomed to the old one, relies on an outdated mental model;

- **Misplaced salience:** The salience is the capability of catching the attention of the users on the important information. This can be generally done by exploiting physical features such as color, blinking lights, sound and others, and it is positive if used correctly, but it can create issues if not. The salient properties help to increase the level of SA if properly exploited, but in cases where too many elements require the operators' attention, it causes the opposite effect;

- **Out-of-the-loop:** One of the most risky demons today is due to the trend of incorporating automation into production processes to replace or support humans. Indeed, automation can support maintaining a high level of SA because it fights the workload demon. However, this leads to the onset of the out-of-the-loop demon: if there is an automatic part, the user tends to rely on it a lot: at this point, if the automatic system fails, it becomes difficult to intervene and even figure out what the problem may be.

2.4 Summary

In this chapter, we introduced the theoretical background on predictive maintenance and maintenance practices in the cultural heritage domain, Cyber-Physical-Social Systems and digital twins, and the fundamentals of situation awareness. In the first part, we traced the evolution of maintenance across different sectors, with particular emphasis on cultural heritage, and discussed how recent technologies have enabled a shift from traditional maintenance toward preventive and predictive maintenance, driven by IoT and AI. We also summarised CH maintenance activities into a dedicated lifecycle, shown in Figure 2.1. We then described CPSS and digital twins as enablers of efficient and effective maintenance, discussing their roles and capabilities in detail. Finally, we introduced situation awareness as a key concept for decision-making in dynamic environments, and analysed Endsley's model and its connections to

decision processes, human and cognitive factors, system design considerations, and common SA *demons*.

Overall, these concepts provide the foundation for the proposal presented in this thesis, namely a Situation-Aware Cyber-Physical-Social System approach for Cultural Heritage Protection: the approach is framed around the Cultural Heritage Maintenance Lifecycle identified in this work, supporting its various phases. Together, these concepts clarify the problems such a system must address and the challenges it is intended to overcome.

Chapter 3

AI Challenges for Cultural Heritage Protection

"There is something irreversible about acquiring knowledge; and the simulation of the search for it differs in a most profound way from the reality."

— J. Robert Oppenheimer, *Physics in the Contemporary World*

In this chapter, we review the state-of-the-art on the use of Artificial Intelligence for Cultural Heritage protection.

Using a detailed bibliometric approach, we analyse current research directions and offer perspectives for future developments in AI for CH restoration and conservation. Specifically, we examine research trends by analysing the most frequently used keywords, authors' nationalities, journals, and other relevant indicators.

However, the main aim of this state-of-the-art chapter is to provide a comprehensive review of AI approaches for CH conservation. We propose a classification of reported approaches based on the types of assets, the data employed, and their scope. Within this framework, we discuss the limitations of existing methods to position our contribution and clarify the advances we introduce.

We observe that real-time monitoring and automated management, enabled by innovative monitoring approaches and devices, can simplify the management of diverse CHCIs, reduce cognitive workload, and broaden the data available to understand ongoing conditions [60]. AI and ML make it possible to analyse such data [61]. Aerial and ground cobots, together with advanced imaging technologies, are reshaping surveying and inspection by providing both synoptic and detailed views through high-quality imagery [9], [62]. Decision-support systems can be integrated to deliver informed recommendations during maintenance procedures [63]. Sustainability and energy efficiency are also increasingly important: new methodologies and innovative techniques not only

improve the quality of maintenance activities but also promote sustainable conservation, enhancing energy performance and limiting environmental impact.

Advanced techniques, such as CNNs and related AI methods, enable the detection and assessment of critical features in artworks, buildings, and cultural artefacts [64]. They support the early identification of deterioration and other issues. AI models can also support restoration by predicting the outcomes of different preservation techniques or recommending suitable approaches based on analyses of similar cases. AI-based analyses of pigments, materials, and artistic techniques can provide insights into historical and cultural contexts. Some studies are also exploring climate-neutral approaches to CH preservation, an increasingly important priority as climate change threatens many historic sites and artefacts.

The following sections present both a trend analysis and a technical review of the identified approaches.

Parts of this chapter have been previously published in:

- Francesco Colace, Rosario Gaeta, Angelo Lorusso, Michele Pellegrino, Domenico Santaniello. *New AI challenges for cultural heritage protection: A general overview*. Journal of Cultural Heritage. 2025. Volume 75. https://doi.org/10.1016/j.culher.2025.07.019.
- Mario Casillo, Francesco Colace, Rosario Gaeta, Angelo Lorusso, Michele Pellegrino. *Artificial Intelligence in Archaeological Site Conservation: Trends, Challenges, and Future Directions*. Journal of Computer Applications in Archaeology. 2025. Volume 8. https://doi.org/10.5334/jcaa.207.

3.1 Methodology

The methodological approach used in this chapter to review the literature on the use of AI for cultural heritage protection follows the steps outlined below and illustrated in Figure 3.1:

1. **Research:** Scopus was used as the main search engine and database, together with IEEEXplore, ScienceDirect, and SpringerLink, to identify indexed papers. We developed a query using relevant and representative keywords and phrases;

2. **Selection:** from the initially identified papers, we selected only those

relevant to our scope, considering both journal articles and conference papers;

3. **Trends Analysis:** trends in the literature were analysed, including the countries and universities of origin, as well as the most frequently used keywords. The trend analysis was conducted using VOSviewer (`https://www.vosviewer.com/`), a tool for bibliometric analysis and for generating citation and collaboration networks;

4. **Technical Analysis:** We organised and analysed the papers according to several taxonomies, covering the types of assets studied, the data employed, and the scope of the proposed approaches. We then discuss the main findings, highlight limitations in the state-of-the-art, and outline possible future directions.

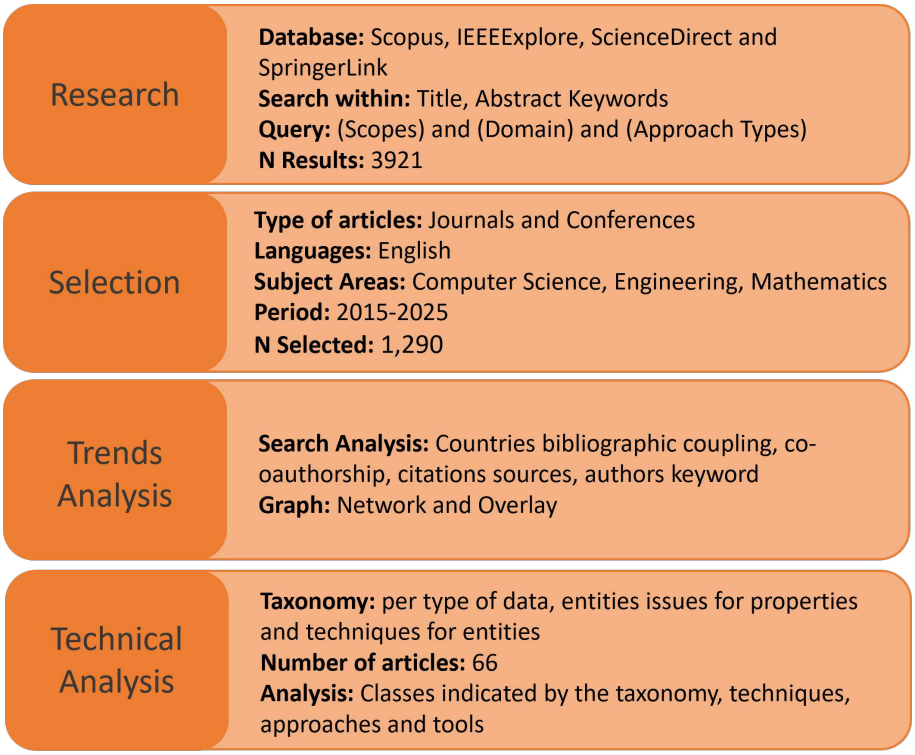


Figure 3.1: Survey Procedure

The documents analysed in this study were retrieved using the following query, which was structured around the research scope, the application domain, and the main classes of approaches and issues addressed. These sections are combined using the *AND* operator.

((predictive AND maintenance) OR (pm) OR (preventive AND maintenance)OR (decision AND support AND system)OR (dss)OR (restoration) OR (reconstruction) OR (preservation) OR (safeguard) OR (protection))

AND

((cultural AND heritage) OR (ch) OR (archaeological AND heritage) OR (historical AND heritage))

AND

((deep AND learning) OR (machine AND learning) OR (artificial AND intelligence) OR (ai)OR (classification)OR (detection)OR (regression)OR (detection) OR (identification))

The first part of the query defines the scope of the search, focusing on the preservation and safeguarding of CH. The second specifies CH as the domain of interest, while the third sets the focus on DL and ML approaches and techniques, as well as on the problems to be addressed.

The results from the initial search were filtered to include only English-language journal and conference papers in the areas of computer science, engineering, and mathematics. The publication window considered is 2015-2025. In particular, the study is updated through June 2025.

A further filtering step retained only papers strictly related to CH safeguarding and protection, while excluding works focused on AI for CH cataloguing and promotion. We also excluded papers that used CH only as a buzzword and did not report case studies directly related to such domain. The 66 selected papers were then used to build the taxonomies presented in Figures 3.5, 3.6, 3.7, 3.8, and 3.9. We also performed a bibliometric analysis using VOSviewer, examining countries' bibliographic coupling, co-authorship networks, and authors' keywords through network and overlay visualisations.

This bibliometric analysis allowed us to identify the countries most active in AI for CH and the keywords that are shaping research trends and scientific discovery in this field.

3.2 Trend Analysis

Before conducting a focused technical analysis of AI-related papers on CH protection, we identify broader trends, including the countries most active in this area, collaborations among them, and the most frequent keywords and journals.

In particular, the bibliographic coupling network shown in Figure 3.2 suggests that the countries most engaged with this topic are those for which CH

protection is strategically important, both culturally and economically, and it also highlights how these countries collaborate to address shared challenges. In this context, Italy and China emerge as the most prolific countries in terms of publications, followed by Greece, India, the USA, and Spain.

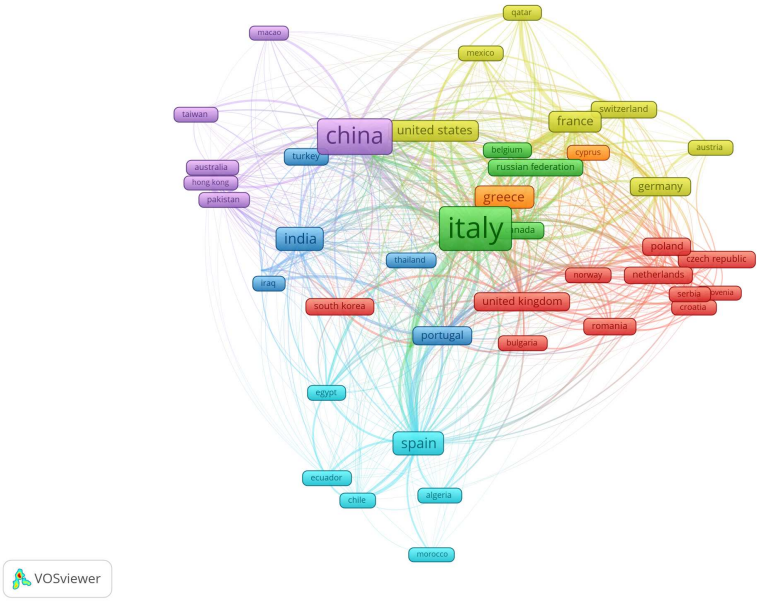


Figure 3.2: Bibliographic Coupling Network

The most active journals and conferences were also analysed, considering only sources that published at least 20 works. The distribution is shown in Figure 3.3. From this analysis, we observe that the Journal of Cultural Heritage, Sensors, and Sustainability are the most productive outlets. We also found a substantial number of works published in IEEE conference proceedings.

The most informative bibliometric analysis concerns the authors' keywords, as they reflect the topics and research areas most central to the field.

Keywords provide an immediate characterisation of literature trends in terms of issues, goals, and approaches, as shown in Figure 3.4.

The most recurring keywords are CH and DL. In this context, AI approaches are often applied to structural health monitoring, where historic buildings represent critical assets that must be safeguarded. Structural health monitoring typically involves identifying damage and vulnerability in buildings, including seismic and climate-related vulnerabilities. LiDAR and laser scanning are frequently adopted technologies, and digital reconstruction is one of the field's main objectives.

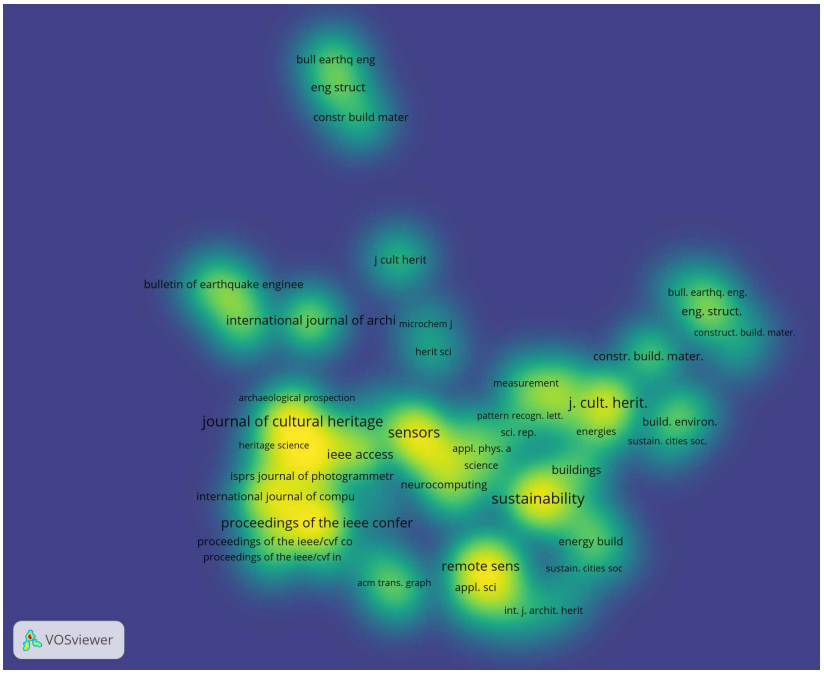


Figure 3.3: Citation Sources Density Map

Among the most frequent terms, BIM and HBIM stand out, as they are often used as a foundation for digital twins and CH preservation [65] [66]. UAVs are also key devices for monitoring CH assets, especially when combined with preprocessing tools and AI techniques. Finally, semantic web technologies and ontologies can be used to represent the CH domain, often starting from existing databases and semi-automatically converting them into ontologies.

3.3 Technical Analysis

In this technical analysis, we define a taxonomy of the different types of CH that can be considered in the CH domain.

This taxonomy is based on two macro-categories:

- **Movable Heritage:** refers to any asset that can be moved from one location to another without altering its essence or causing damage to the item itself or to the context from which it is removed.
- **Immovable Heritage:** refers to assets that are fixed to a specific location and cannot be moved without altering their fundamental characteristics or causing damage.

Figure 3.4: Authors Keywords Density Graph

Building on this, we propose a comprehensive taxonomy of CH inspired by UNESCO's classification of CH [67], [68]. This taxonomy is shown in Figure 3.5.

In particular, we identified papers focused on the protection of sculptures, manuscripts, and paintings. We also found a variety of techniques for detecting issues and supporting the preservation of different types of paintings, including canvas works, mosaics, prehistoric paintings, and frescoes.

Canvas paintings can generally be classified as movable heritage, whereas prehistoric cave paintings and frescoes are typically considered immovable, as they are integrated into built structures or natural surfaces. For mosaics, the classification depends on the context, but most of the literature addresses issues related to mosaics that are part of immovable heritage, such as those found within buildings and archaeological sites.

The literature analysis identified approaches and techniques focused on five types of immovable heritage:

- **Archaeological sites**, like ancient settlements, funerary areas, and caves;
- **historical buildings**, such as public monuments, churches, and castles;
- **historical urban areas**, such as ancient hamlets or squares;
- **logistic infrastructures**, such as bridges;
- **museums**, and other exhibition and dissemination venues.

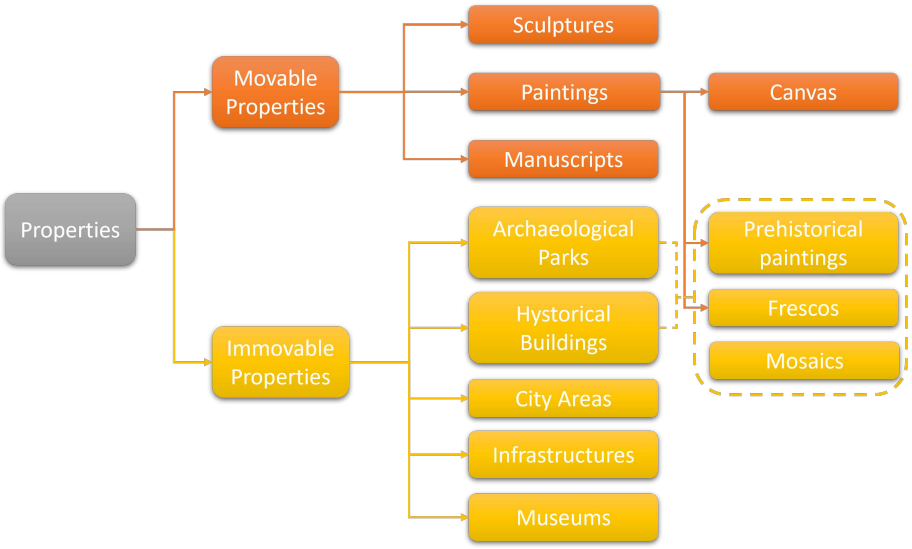


Figure 3.5: General Taxonomy of Heritage and Assets

Prehistoric paintings and frescoes are generally considered immovable, as they are typically found in archaeological sites and historic buildings. We also identified a smaller number of works that apply AI techniques to CH preservation using historical photographs and videos through rephotography.

In the following subsections, we outline the goals and data associated with the subcategories of movable and immovable heritage.

The description follows the classification shown in Figure 3.5, which distinguishes between movable and immovable heritage.

The goals reported in Figures 3.7 and 3.9 were derived by reviewing the full set of selected papers and analysing the works discussed in the following sections. This process allowed us to identify the goals addressed by each paper and to summarise the current state of the literature. The same procedure was used to identify the data types reported in Figures 3.6 and 3.8.

3.3.1 Movable Heritage

The first category of papers analysed in the literature concerns the safeguarding of movable heritage. We derived the taxonomies shown in Figures 3.6 and 3.7, which report, respectively, the types of data used by the reviewed approaches and the main goals and issues addressed in the literature. These taxonomies show that many works focus on damage detection, such as identifying mildew on canvas paintings, often using datasets that include multiple predictors collected

by sensors. Semantic segmentation applied to 3D models is also used for detecting damage on sculptures.

Another widely investigated topic is the identification of entities within paintings. This is frequently addressed alongside the restoration of artworks in disrepair to their original condition and the detection of maintenance issues affecting movable heritage.

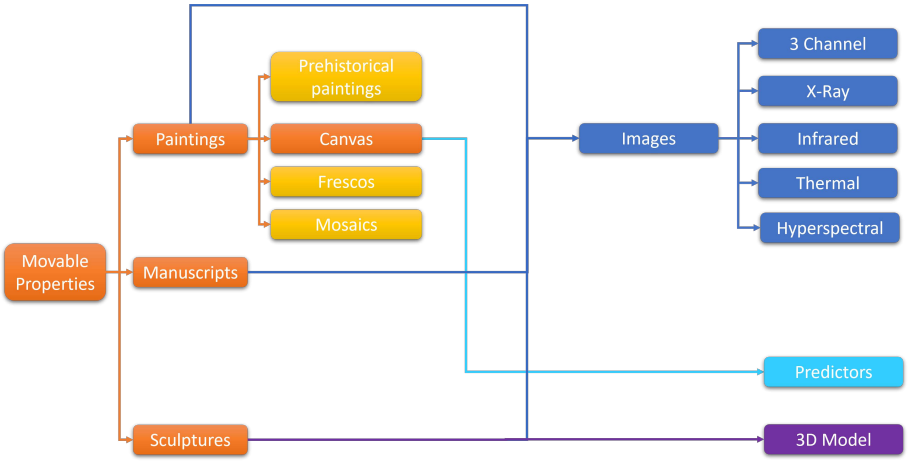


Figure 3.6: Classification of Movable Heritage and Type of Data

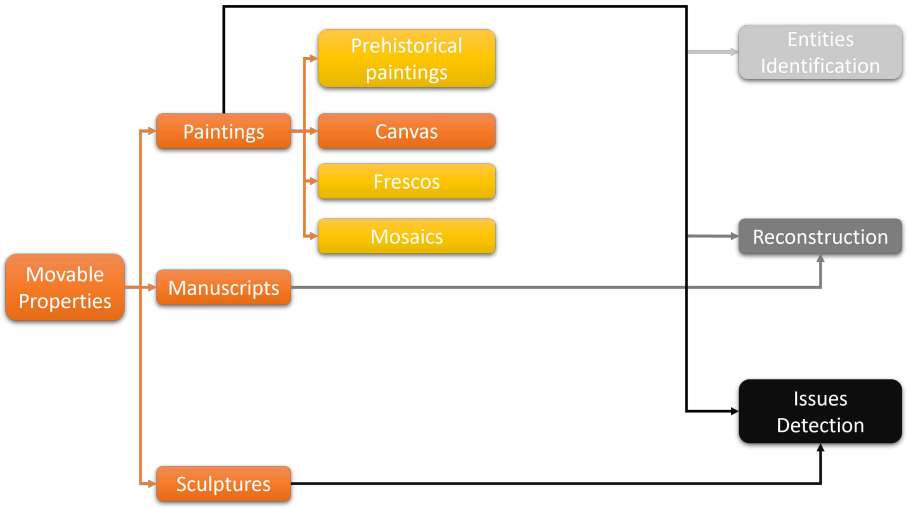


Figure 3.7: Classification of Movable Heritage and Goals

Since in this work we are more interested in AI techniques for the protection of CHCI, generally real estate, we will not go into detail about work related to

movable heritage, but we will list them and briefly discuss what we have found.

3.3.1.1 Overview and Key Findings

Work	Asset	Goal	Type of Data	Technique	Result
[69]	Paintings	Issues Det	Images: RGB, Infr and X-ray	Convolutional MCNC	Detection: 0.8185 F1-m
[70]	Paintings	Issues Det	Images: RGB	Res-UNet	Segmentation: 93.42% F1-m
[71]	Paintings	Issues Det	Images: Ther	Mask R-CNN	Segmentation: 91.04 F1-m
[72]	Paintings	Issues Det	Images: Ther	YOLOX-II-Tiny	Detection: 96.67% mAP
[73]	Paintings	Issues Det	Predictors	Random Forest	Classification: 0.62% A
[74]	Paintings	Issues Det	Predictors	Decision Tree	Classification: About 0.5 AUC
[75]	Paintings	Issues Det	Images: RGB	ResNet	0.89 F1-m
[76]	Paintings	Rec	Images: RGB	Patch-based method + mscale near-neighbour + Gaussian pyramid	NA
[77]	Paintings	Rec	Images: RGB	Pix2Pix GAN (U-Net + PatchGAN)	IoU 0.62
[78]	Paintings	Rec	Images: RGB	Mask R-CNN + U-Net	9.22 MSE, 0.952 SSIM
[79]	Paintings	Rec	Images: RGB	PConv and EdgeConnect, YOLOv5	0.8301 Edge 0.8404 PConv SSIM 0.9530 mAP@0.5
[80]	Paintings	Rec	Images: RGB	Bezier, Laplacian and Gaussian pyramids	0.8760 LPIPS, 0.9494 DISTS
[81]	Paintings	Rec	Images: RGB	DDCNN	NA
[82]	Paintings	Rec	Images: RGB and Raw XRF	Custom CNN 1D	NA
[83]	Paintings	Rec	Images: RGB	TRT and DFP with ROI network + edge detection AOI	0.9268 A, 0.967 SSIM
[84]	Paintings	Rec	Images: RGB	GAN U-Net + Transformer + custom CNN discriminator	76.3% AP, 78.2% A
[85]	Paintings	Rec	Images: RGB	DunHuangStitch	35.07 PSNR, 0.949 SSIM
[86]	Sculpture	Issues Det	Images: Hyp	DBSCAN with PCA	NA
[87]	Sculpture	Issues Det	3D Model	Change-based seg + k-neighbour hist	NA
[88]	Manuscript	Rec	Images: RGB	UNet + Partial Convolution	38.17 PSNR, 0.972 SSIM (16x16 defect)
[89]	Manuscript	Rec	Images: RGB	ViT + DCT	90% F1-Score

Table 3.1: Summary of Approaches and Results on Movable Heritage

The papers on movable heritage are summarised in Table 3.1. For each approach, the table reports the type of asset considered, the paper's goal, the data on which the approach operates, the technique proposed or adopted, and the reported results. Most works focus on paintings, addressing detection, segmentation, or classification tasks to identify issues. Many studies

fine-tune pretrained state-of-the-art networks. For instance, YOLO, a widely adopted object-detection architecture, is frequently used, especially in its v5 and YOLOX variants, while Mask R-CNN is widely used for semantic segmentation. Several segmentation models are U-Net based, often with a ResNet backbone, which is also common in classification tasks. Other works propose custom networks to address detection or segmentation problems, or rely on traditional ML techniques such as RF and DT, using predictor-based datasets.

For damage detection in paintings, a range of image-based techniques is employed, from custom CNNs, such as MCNC for crack detection, to Mask R-CNN for surface and subsurface defect detection using thermal imagery. Some models, such as Res-UNet, report accuracies up to 98% on specific datasets [70]. GANs are often used for painting reconstruction, including Pix2pix and U-Net-based variants. Other DNNs, such as PConv and EdgeConnect, typically operate on RGB images and aim to restore original conditions by addressing issues such as colour degradation and tears, with promising results in terms of SSIM and PSNR [79] [83].

Studies on sculptures mainly focus on issue detection, using hyperspectral imagery combined with methods such as PCA and DBSCAN, as well as 3D models for surface geometric change analysis. For manuscripts, the focus is almost exclusively on reconstruction, with approaches such as U-Net and partial convolution used to repair holes and tears. While reconstruction methods typically start from RGB images, issue-detection approaches often rely on spectral information, including infrared, X-ray, and thermal images.

Reported results vary across studies. Some papers achieve high accuracy, although these results are often influenced by measurement uncertainty and limited data availability. Others report less conclusive outcomes, or omit quantitative results altogether, because they are still at an early experimental stage. They are nevertheless included because they help clarify the dominant research directions. Overall, the choice of data type, ranging from three-channel images to spectral imagery and sensor-derived predictors, is tightly linked to the target objective. Despite encouraging results, the limited availability of training data and the variability across datasets highlight the need for larger and more standardised benchmarks to improve model robustness.

3.3.2 Immovable Heritage

Safeguarding and protecting immovable heritage is also widely addressed in the literature. As shown by the taxonomies in Figures 3.8 and 3.9, studies on immovable heritage protection rely on several data types, with a strong emphasis on imagery. Aerial laser scanning is frequently used to acquire both 2D and

3D data. In addition, various predictors collected by sensors are employed in the context of archaeological sites, historic buildings, and museums.

Alongside entity identification and issue detection, another recurring research direction concerns visitor support, that is, methods designed to assist users during visits to places such as museums. We include these works because they can be valuable for monitoring and managing visitor flows, providing data that contribute to understanding the vulnerability of artifacts and structures. For this reason, they are relevant to CH protection.

The taxonomies also show that museum-related studies often focus on developing methods to monitor visitors and enhance their experience during visits.

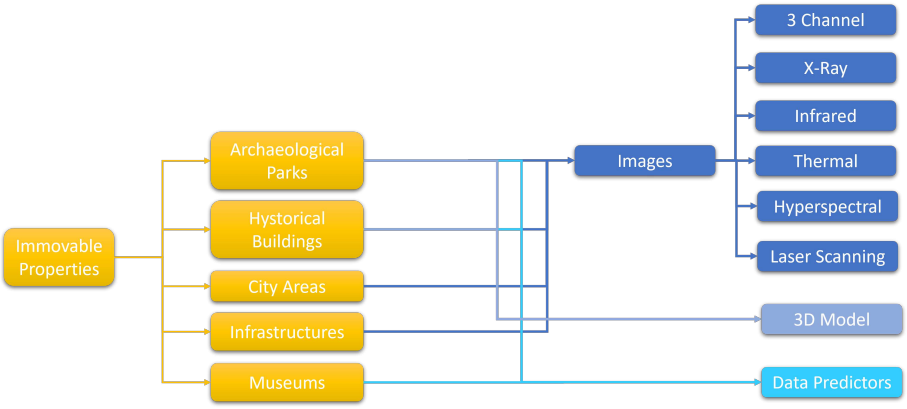


Figure 3.8: Immovable Heritage Classification with Type of Data

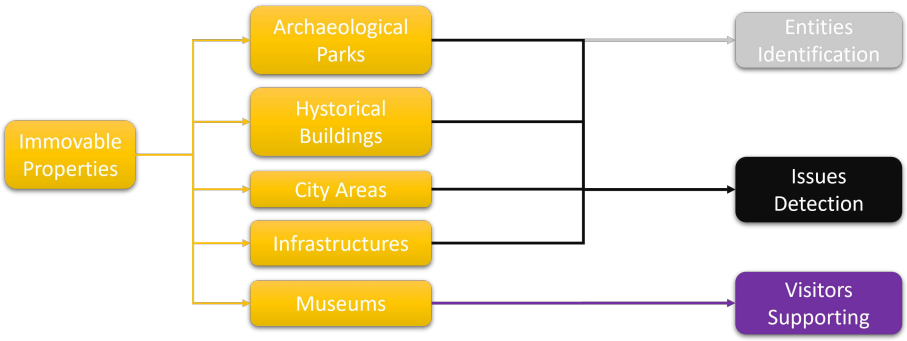


Figure 3.9: Immovable Heritage Classification with Goals

3.3.2.1 Archaeological Sites

In the context of this work, it has been observed that authors are generally very interested in proposing and researching approaches for identifying new sites or entities in existing sites.

Tao et al. [90] propose an approach for the detection of villages in the Conghua area from remote sensing images using deep learning and geographical coordinates. The authors tested two models, one for classification and one for detection. VGG16 was found to be 90.79%. Retinanet for object detection reaches an accuracy of 95.61%. The approach explained in this study can contribute to the detection of ancient settlements, enabling their protection. Altaweel et al. [91] deal with the automatic identification of archaeological features using RGB images coming from UAVs. This model achieves an accuracy of over 0.9. The authors developed an application, with a Graphical User Interface for annotation, using the Mask Region-Based Convolutional Neural Network (Mask R-CNN) for instance segmentation. Trier et al. [92] deal with the mapping of the Norwegian landscape. Faster R-CNN is used to detect features in ALS data. Considering 737 test images (16.6 km^2), 87% of the natural and presumed historical heritage items were assigned to the right class (consumer accuracy). Only 1% of the CH elements were misclassified; 13% of the examples were not assigned. To improve the system, false positives must be reduced for photos covering larger areas or entire mapped areas. The work of Anttiroiko et al. [93] proposes a method for detecting tar production kilns, which is a phenomenon in the northern boreal forest of Finland. The U-Net module was used to classify the archaeological traces of the tar furnaces from the ALS data. The proposed system showed high efficiency, with an overall accuracy of 93% in detecting the kilns. Li et al. [94] use an improved YOLOv8 network combined with multi-scale attention module (EMA) for dense detection and counting of Buddha head sculptures in the Wanfo Hall in Beijing's Fangshan district. The proposed approach achieved a mean Average Precision (mAP) of 0.92, handling densely arranged and partially damaged Buddha heads. Bhuvaneswari et al. [95] use a deep learning-based system for recognizing and analyzing ancient Tamil temple inscriptions. Various preprocessing activities are performed and combined with Vision Transformers with multi-head attention and transfer learning. The model achieved high performance: 97.25% precision, 95.05% recall, 95.17% F1-score, and 98.92% accuracy. A key challenge remains the variability in script styles and noise due to erosion.

Issues detection is another fundamental problem addressed in literature. Zhang et al. [96] find fractures in the façades of clay monuments using DL. The team created a DL algorithm to identify single pixel level fractures. Three DL systems based on U-Net, Linknet, and FPN with distinct architectures were implemented. A custom dataset of cracks from historical sites was collected and

used in combination with existing damage image collections. The FPN-vgg16 approach achieved the highest F1 score (84.40%) and IoU score (73.11%), indicating significant potential for application in crack identification in earthwork sites. The main drawback of such approach was the difficulty of implementing these systems in extremely light-embedded devices. Altaweel [97] deals with deep learning for the monitoring of looting phenomena at CH sites: the model is based on the Mask R-CNN architecture, trained on a dataset of aerial images collected in the Middle East, reporting various types of damage and changes caused by looting. The research showed that the method is highly capable of identifying areas of looting, achieving an accuracy of 93%. The approach sometimes fails in the detection of corroded damage. Qiu et al. [98] detect damaged roofs in traditional Chinese residential buildings with a YOLOv8-based method. Six villages in Xiamen and Quanzhou are scanned using a UAV, creating a dataset of 3124 images. The process involves first selecting historically valuable buildings, with a 97.2% mAP and then segmenting roof damage, with a 89.4% mAP. The authors highlight limitations such as dataset variability and insufficient coverage of various roof damage types. Peng et al. [99] introduce a deep learning-based radial basis function (RBF) for compressing hyperspectral images in the zone of the Baoding Mountain Grottoes, to find deterioration in stone cultural relics. The approach significantly enhances identification efficiency, reducing training and detection times by approximately 4.7% and 9.1%, respectively, while maintaining an F1-score around 0.95. Corradetti et al. [100] use LLM approaches like GPT-4o, Gemini 1.5 Pro, and Claude 3 Opus for identifying stone deterioration patterns in CH. The dataset used is obtained from archives of global heritage sites. GPT-4o achieved the highest recall (42.1%) in identifying the main "target" pathology, with Gemini close (39%) and Claude trailing (24.3%). Precision for recognizing additional non-target patterns was higher, with Gemini scoring best (69.5%). Such models showed inconsistent results and semantic confusion due to lack of samples. The RaViT-AE proposed by Kwon et al. [101] is a ViT-based autoencoder for unsupervised anomaly detection in CH monitoring, used to target petroglyph images from the Bangudae Terrace in Korea. The method uses region-attentive patch embedding, and proposes a F-SSIM loss, which combines SSIM and high-level feature loss on ResNet152. RaViT-AE achieved strong results, in particular AUC: 0.976, Accuracy: 0.944 and F1: 0.936, outperforming GAN and CNN-based models. The method struggles with subtle anomaly detection and performance under extreme conditions.

The use of LIDAR (Light Detection and Ranging) technology and DL techniques to identify ancient Maya archaeological sites is discussed in Rissetto et al. [102]. The PointConv architecture is proposed to catalog Maya archaeological sites from 3D LIDAR data in point clouds. PointConv showed greater accuracy than the CNN-based approach used for comparison, with 95% accu-

racy with the model trained with data augmentation.

Other sensor data can also be used in such a context, for example by Mertel et al. [103], who use a graph similarity method for predicting prehistoric sites in the Bohemian-Moravian highland. Three main steps are executed: the association of archaeological details with the environmental characteristics of each site, the comparison of individual sites using Hamming distance, and the prediction of the presence of specific archaeological components in areas where there are no finds. The method provides an AUC of 0.62. Wachtel et al. [104] compare two models, based on LoR and graph similarity analysis. The LoR method required the availability of data about the archaeological sites, which is difficult due to the frequent inaccuracies in archaeological data. The MaxEnt method proves to be more flexible and more effective in cases when few archaeological site observations are available. They achieved an AUC of 0.796 ± 0.02 with 80% training and 20% testing. Yaworsky et al. [105] use an ML approach for finding archaeological sites in the Grand Staircase-Escalante National Monument. GLM, GAM, MaxEnt, and RF approaches are tried. Each method was evaluated using performance measures such as AUC and actual ability statistics (TSS). The MaxEnt approach, which uses pseudo-absence points, has proven to be the most effective for these data types, obtaining an AUC of 0.88. Imen et al. [106] apply a MaxEnt self-learning method to locate urban heritage sites in the Saoura region of Algeria. They investigate which features among slope, elevation (DEM), distance to water, normalized difference vegetation index (NDVI), fertility, and proximity to palm groves, mostly influence the sites' location. Fertility is the most influential component, accounting for 94.1% of the total weight. The MaxEnt module achieves an AUC of 0.859.

3.3.2.2 Historical Buildings

The protection of Historical Buildings is probably one of the most addressed issues in such a context. Zou et al. [107] aim to detect the position of missing components in historic buildings, which is a critical aspect of CH conservation. The technology involves training the Faster R-CNN module using images collected from historic buildings, specifically the Forbidden City in China. They obtained an AP of 0.8831. Liu et al. [108] aim to identify cracks in wood inside historical heritage sites. They use the YOLOv3 algorithm with the DarkNet-53 network as the backbone for crack detection, which showed high accuracy, exceeding 90 percent. The inference time of the approach is 0.1 seconds. Wang et al. [109] propose a system to segment and measure damage on the glazed tiles of historic buildings. Faster R-CNN is used to distinguish between two types of glazed tiles. The module achieved an AP of 0.910 and 0.890 for the first two

types of cropped images, respectively. The cropped images were then used as a dataset to train the Mask R-CNN model afterward. The Mask R-CNN module is used to perform segmentation and damage measurement on the cropped images. In this case, the method can accurately segment the damage from the pixel level, achieving an AP of 0.975 for damage separation. Mask R-CNN is shown to be superior to an FCN. Mishra et al. [110] use an AI visual inspection method to detect flaking, exposed walls, cracks, and detachments. They compare YOLOv5 and Faster R-CNN. The results indicate that the YOLOv5 model outperforms the Faster R-CNN model in terms of mAP, achieving a more accurate defect detection with a mAP of 93.7 percent. Mansuri et al. [111] aim to perform automatic inspections in an architectural heritage context. Flakes, exposed brickwork, and cracks are automatically identified using Faster R-CNN. The approach requires a large dataset of high-quality images, and the adaptation to different types of structures can be difficult. The mAP achieved is 0.915, with superior performance in terms of time and resources required. Elhariri et al. [112] focus on pixel-level cracks of historical surfaces. Variants of U-Net, named Deep ResU-Net, ResU-Net++, and U2-Net, are proposed for the detection of cracks in old town and historical buildings. The U2-Net model proves to be the best, achieving Dice scores and mIoU of up to 71.09% and 78.38%, respectively. The method must be trained with a large annotated dataset, and it is expensive. Samhouri et al. [113] use two CNN models in cascade to determine whether damage is present and its type. In such a context, VGG16 is used for feature extraction and classification. High-resolution color images were used to train and examine the two models. The approach shows an accuracy of 95 percent in detecting damages. The system's reliance on high-quality images, as well as its lack of adaptability, are limiting factors. Lee et al. [114] aim to detect defects and assess the security of immovable cultural property using a Deep Learning framework. EfficientNetB0, EfficientNetB2, ShuffleNetv2, and AlexNet are used and compared. Results show a high rate of accuracy in the diagnosis of damage: 99.69% for EfficientNetB0, 99.66% for EfficientNetB2, and 93.46% for ShuffleNetv2. Bruno et al. [115] use RGB images for recognizing and segmenting deterioration in historic buildings. Mask R-CNN is used to identify the deterioration patterns of architectural elements in historic buildings. Images of castles, monasteries, and noble and rural buildings with wet spots and biological colonies are used for training and testing. The average accuracy over three images is 1, and the IoU is 0.77, 0.62, and 0.69, respectively. However, the approach fails to accurately recognize the presence of biological colonization. Kwon et al. [116] aim to find damage in stone CH using a Faster R-CNN neural network method to classify and place damage in photographs. The new method achieved an average CR of 94.6% on 20 test photographs. The approach needs improvement for small damaged areas and poorly defined edges. Choi et al. [117] propose F-AnoGan to find damage

and displacement of wooden cultural property through anomaly detection. The model was trained on a custom dataset and computed an anomaly value to identify damaged or displaced areas in the tested photos. Despite its efficiency in differentiating natural images from outliers, the method has difficulty detecting minor irregularities in outlier photos. Lee et al. [118] deal with the automatic classification of wood surface defects using DL algorithms based on RGB images. Damage such as cracks, exfoliation, and decay must be detected using non-invasive methods. The research compared EfficientNetB0, VGG16, ResNet50 and InceptionNetV3. EfficientNetB0 had the highest accuracy, with 96.50 percent. Roy et al. [119] identify drips, cracks, and colour variations in paint by using CNNs and SVMs. The CNN-SVM showed high accuracy, with an overall accuracy of 84.23%. Sohail et al. [120] use Faster R-CNN combined with a fuzzy rule system for identifying thick walls, moats, towers, and drawbridges. The method achieves an accuracy of 94.16% in recognizing castle towers. Mehta et al. [121] combine CNNs and SVMs to classify the danger of historic buildings. They classify damage as no damage, light, medium, and high damage. CNN-SVM model outperforms the SVM model in terms of accuracy, recall, and F1 score, achieving an accuracy of 85.6%. Conversely, the SVM model attained an accuracy of 80.2%. Meklati et al. [122] use a mobile application to take and upload photographs of any suspected damage, analyzing with MobileNetV2 CNN. The approach obtained good results for weighted average accuracy (0.868) and average recall (0.84). Yu et al. [123] use thermal images to identify defects in standard brick, asphalt and red brick of historic buildings using AlexNet. The method shows a 92.1 percent identification rate, a relevant advantage compared to a more time-consuming manual type of analysis. Gao et al. [124] detect fire events in historic buildings using video. In the first stage of their methodology, a Gaussian mixture model is used to identify the elements in action. Fire-Det, which is able to detect fire effects in real-time, uses EfficientNetV2 feature extraction and is based on a spatial pyramidal pooling layer based on Spp-net for feature fusion and the YOLOV3 structure for detection. The model obtains a precision of 78.1 and a mAP of 76.1. Bakirman et al. [125] use DexiNed and RCF, two deep learning architectures, to extract profiles and cracks from CH structures. The study produced F1 scores of 61.38% and 61.50% for the two architectures. Wang et al. [126] are the authors of GreatWatcher, a system designed for rapid investigation and defect detection in CH preservation. Using such a system, mobile users collect photos and data that are sent to the web platform and are processed by a ZF-Net based approach, by replacing the pooling layers with a pooling layer of ROI, for automatic damage detection. Thus, the network achieved an accuracy of 78.2 percent. Lorusso et al. [127] use GANs as a graphical support tool for historic urban renovation. The CycleGAN approach, which involves two generators, and two discriminators connected to two Forward and Backward cycles,

generating two separate loss functions was applied to find the characteristic colour style in historical building facades. Cheng et al. [128] propose DeblurDetNet that detects defects such as cracks, corrosion, biological invasion, and others, in historical buildings, addressing image blurriness. It has been tested on historical buildings in Guangzhou, China. It achieves a mAP of 0.553 after image enhancement. Singh et al. [129] propose a YOLOv8-based method for detection of cracks, spalling, and vegetation growth at Darbhanga Fort in India. The YOLOv8 model achieves a mAP of 96.2% for detection and 92.5% for segmentation. The authors compare YOLOv8 to Faster R-CNN and YOLOv5. Long et al. [130] introduce MSD-Det, a large-scale, high-resolution dataset for detecting issues such as cracks, unit loss, surface dissolution, efflorescence, discoloration, plant growth, and moss in masonry structures. Seventeen deep-learning object detection algorithms were evaluated, such as YOLOv5, YOLOv6, YOLOX and Faster R-CNN, with YOLOv5-x achieving the highest performance, in particular mAP@0.5 of 93.2%.

Considering the possibility of working with 3D models, IDJATON et al. [131] find limestone deterioration in 3D relief images using YOLOv5 and transformation layers. They achieved an F1 score of 85% and an average accuracy of 81%, outperforming existing approaches. The approach often fails to detect small areas of flaking and struggles to detect them reliably. The paper [132] uses Deng entropy combined with a bidirectional Long Short-Term Memory (bLSTM) neural network for classifying severely damaged buildings. It uses voxel representations derived from point clouds to quantify structural complexity and damage. The proposed method achieves perfect accuracy and Matthews correlation coefficient (MCC) scores on validation tests.

Many studies are also related to the use of predictors obtained from sensor data, as Zonno et al. [133] focus on the analysis of the long and short-term effect of temperature and humidity on the Church of San Pedro Apostol in Andahuaylillas, Peru. The methodology uses aerial and terrestrial photogrammetry, terrestrial laser scanners, and dynamic identification testing to describe such buildings. An MLRM was developed to understand relationships between natural frequencies and environmental parameters, achieving relative errors of less than 3-4%. Miglioranza et al. [134] deal with the damage caused by indoor microclimate fluctuations due to human presence, with a case study in Ringebu and Heddal stave churches in Norway. The DPC used aims to detect patterns in data distribution. CNN classifies days using the internal mixing ratio (MRin) time series. The UMR peak works on the heuristic assumption that days with scheduled events have at least one peak in the MRin time series that is not explained by the external (MRout). The results show that the UMR peak outperforms previous methods regarding accuracy, recall, and NMI. Finally, the UMR reaches 0.73 of precision, 0.82 of recall, and 0.53 of NMI. The work by Carrara et al. [135] is based on attentional neural networks. The model

is trained with data collected in a long-term experimental phase, including sensor detections and environmental data. The model is then used to predict the structure's natural frequencies using this data. Anomalies were detected by comparing deviations between actual and predicted frequencies. Colace et al. [65] propose an IoT-BIM-based framework capable of improving the enjoyment and protection of CH objects. The inference engine of such system uses the CDT contextual filter to provide recommendations. The CARS engine enables the personalization of user preferences and the selection of appropriate POIs. The project aims to evaluate the accuracy of the CARS and PM modules. The same authors propose a different approach in [136] using data from a sensor network but a different framework, based on Analytic Network Process (ANP) with Zero-One Goal Programming (ZOGP) for the decision strategy. Casillo et al. [137] focus on predictive preservation, which includes adopting behaviors to reduce material loss and damage to tangible CH. The system is managed via a DTW constructed on an HBIM model and integrated with the cloud-based data management platform, named ThingsBoard. Casillo et al. [138] propose an IoT-based framework integrated with Machine Learning techniques for the preservation of historical buildings. The collected real-time data from the real case study of the Scientific Library of the University of Salerno feed AutoEncoders for anomaly detection and KNNs classifiers to recommend interventions based on detected damage. The system achieved accuracy above 90% in anomaly detection. Nota et al. [139] combine the situational awareness and cyber-physical system concept to manage heritage buildings, using the Royal Palace of Carditello as a case study. The framework employs edge-cloud architecture, smart objects with IoT sensors, and machine learning algorithms to continuously assess building conditions. The case study demonstrates effective detection and response to environmental and structural anomalies.

3.3.2.3 Urban Areas

This section addresses AI approaches for protecting urban areas, such as squares or historic districts in modern cities.

The paper presented in [140] analyses the possibility of using AI models for tidal level prediction in Venice, focusing on high water. The proposed approach is trained on historical tidal data, considering variables such as astronomical tide, wind speed and direction, barometric pressure, and past tidal levels observed in the past: this enables tidal level prediction, particularly during high water events. The approach requires a large and high-quality dataset for effective model training. The M5P algorithm-based model generally outperforms Random Forest and Multilayer Perceptron models in predicting tidal levels. The model achieves an R2 of 0.9958 and a RAE of 5.98 percent. These models

proved particularly effective in predicting high tides and demonstrated good accuracy in forecasting high water events.

3.3.2.4 Infrastructures

In the CH context, safeguarding infrastructures often focuses on preserving historic bridges made of various materials.

Cardellicchio et al. [141] combine CNNs and XAI to identify and assess damage to existing reinforced concrete bridges. MobileNetV3Small demonstrated the best performance in finding defects such as cracks, corrosion, and delamination, achieving an accuracy of 63.46%, a precision of 69.97%, and a recall of 55.42%. Grad-CAM is used that shows which parts of the images were fundamental for network decisions. Lee et al. [142] aim to protect cultural and national heritage and propose a case study on the Woljeong Bridge in South Korea. IoT sensors were used to monitor the bridge, acquiring data like slope, load, and distortion. A wireless ZigBee environment is used for the configuration, including sensing and communication units. An RNN with LSTMs and GRUs and hybrid models are used.

3.3.2.5 Museums

In the literature, many papers addressing the safeguarding of museums focus on supporting visitors during their tours and within different exhibition halls.

In Piccialli et al. [143] LSTM Autoencoder and an IoT control network are used together to predict visitor paths in the context of CH. Subsequence Accuracy, which measures the proportion of correctly predicted bigram sequences, and Presence Accuracy, which assesses the proportion of predicted nodes are used for evaluation. The results are Subsequence Accuracy of 75.65% and Presence Accuracy of 86.89%. Martella et al. [144] want to understand the orient mechanisms of visitors in museum spaces and their interest in art pieces. Proximity sensors are used to detect the location and movements of visitors. Prediction techniques are also employed to forecast visitors' future paths. Based on an R-matrix of 182 x 45 evaluations, the results obtained through the non-negative matrix and non-negative matrix factorization achieve an RMSE of 1.1 and 2.4 for prediction.

3.3.2.6 Synthesis and Key Findings

The approaches discussed for immovable heritage are summarised in Table 3.2. As for movable properties, the table reports the type of asset considered, the paper's goal, the data on which the approach operates, the technique proposed

or adopted, and the reported results.

Many papers on archaeological sites focus on identifying something. In these cases, authors typically work with aerial imagery, acquired via UAVs or satellites, or with data derived from laser scanning. Mask R-CNN and U-Net-based models are widely used for segmentation, while Faster R-CNN and YOLO-based networks, such as YOLOv8, YOLOv7, and YOLOv5l, are commonly adopted for object and issue detection. For example, YOLOv8 combined with a multi-scale attention module achieved an mAP of 0.92 for Buddha head detection, and Mask R-CNN reported 93% accuracy in looting monitoring. In this area, we observed that relatively few works focus on decision-support components.

Studies on historic building safeguarding often rely on RGB image datasets and focus on issue detection. Faster R-CNN and YOLO-based models are frequently adopted, with reported mAP values up to 0.93. Most segmentation approaches are based on Mask R-CNN and U-Net variants, including autoencoder-based models such as U2-Net, which has been used for pixel-level crack segmentation, reaching an mIoU of 78.38%. Classification is typically performed using architectures such as VGG16, EfficientNetB0, and MobileNet. Other works propose custom networks and combined approaches that integrate deep-learning and non-deep-learning methods.

Issue detection is also explored for infrastructures, for instance using MobileNetV3Small, with reported accuracies of 63.46% for defect identification in images. LSTM networks are used to predict future values with high accuracy from sensor-derived predictors. Similar LSTM-based approaches have also been applied in museum settings to predict paths and analyse visitor behaviour.

Overall, the choice of data type, ranging from RGB images to spectral imagery (infrared, X-ray, and thermal), 3D data from aerial laser scanning, and sensor predictors, is crucial for the target objective. Despite variability across studies, reported metrics such as accuracy, AUC, F1-score, AP, and mAP suggest that current approaches are generally promising for issue detection in archaeological sites and historic buildings.

Few projects focus on developing support systems for human operators, even though operators remain fundamental in the CH domain. Since they make the final decisions, any findings must be presented in the most useful and actionable way possible. Recommendation systems and Digital Twin-based approaches partially address this need, but explicit decision guidance and support for operators' situation awareness are still limited. When these aspects are overlooked, operators may be provided with relevant information yet still struggle to interpret it and make timely, well-informed decisions.

Table 3.2: Summary of Approaches and Results on Immovable Heritage

Work	Object	Goal	Type of Data	Technique	Result
[90]	Sites	Iden	Images: RGB	Classification: VGG16, Detection: Resnet	Class: 90.79% A Detect: 95.61% A
[91]	Sites	Iden	Images: RGB	Mask R-CNN instance segmentation	Over 0.9 A
[92]	Sites	Iden	Images: laser scan	Faster R-CNN Detection	87% correct class Less 1% wrong class 13% not detected
[93]	Sites	Iden	Images: laser scan	U-Net based semantic segmentation	93% A
[94]	Sites	Iden	Images: RGB	YOLOv8 network with EMA	0.92 mAP
[95]	Sites	Iden	Images: RGB	Vision Transformers with multi-head attention and transfer learning	95.17% F1
[102]	Sites	Iden	2D and 3D data	PointConv Detection	95% A
[103]	Sites	Iden	Predictors	Graph analysis and comparison using hamming distance	0.65 AUC
[104]	Sites	Iden	Predictors	Max Ent	0.796 ± 0.02 AUC on control group 14 and 20% test
[105]	Sites	Iden	Predictors	Max Ent	0.88 AUC
[106]	Sites	Iden	Predictors	Max Ent	0.859 AUC
[96]	Sites	Issues	Images: RGB	FPN-vgg16 Detection	84.40% F1-m 73.11% IoU-s
[97]	Sites	Issues	Images: RGB	Mask R-CNN Segmentation	93% A
[98]	Sites	Issues	Images: RGB	YOLOv8-based	historically valuable buildings 97.2% mAP roof damage segmentation 89.4% mAP
[99]	Sites	Issues	Hyperspectral Images	Deep learning-based RBF	0.95 F1
[100]	Sites	Issues	Images: RGB	GPT-4 Gemini 1.5 Pro	GPT best R 42.1% Gemini best P 69.5%
[101]	Sites	Issues	Images: RGB	RaViT-AE	0.936 F1
[107]	Buildings	Issues	Images: RGB	Faster R-CNN Detection	0.8831 A
[108]	Buildings	Issues	Images: RGB	YOLOv3 DarkNet-53 based Detection	up 90% A
[109]	Buildings	Issues	Images: RGB	Mask R-CNN Segmentation	0.975 AP
[110]	Buildings	Issues	Images: RGB	YOLOV5 Detection	93.7% mAP
[111]	Buildings	Issues	Images: RGB	Faster R-CNN Detection	0.915 mAP
[112]	Buildings	Issues	Images: RGB	U2-Net model for segmentation	71.09% Dice score 78.38% mIoU
[113]	Buildings	Issues	Images: RGB	VGG16 based classification model	95% A
[114]	Buildings	Issues	Images: RGB	EfficentNetB0 based classification model	99.69% A
[115]	Buildings	Issues	Images: RGB	Mask R-CNN Segmentation	0.98-0.99 OC AP 1
[116]	Buildings	Issues	Images: RGB	Faster R-CNN Detection	94.6 CS
[117]	Buildings	Issues	Images: RGB	F-AnoGan Anomaly Detection	NA
[118]	Buildings	Issues	Images: RGB	EfficientNetB0 based Detection	96.50% A

Continued on next page

Table 3.2 – Continued from previous page

Work	Object	Goal	Type of Data	Technique	Result
[119]	Buildings	Issues	Images: RGB	CNN-SVM Classification	84.23% A
[120]	Buildings	Issues	Images: RGB	Faster R-CNN Detection	94.16% A
[121]	Buildings	Issues	Images: RGB	CNN-SVM Classification	85.6% A
[122]	Buildings	Issues	Images: RGB	MobileNetV2 Classification	0.868 P 0.84 R
[123]	Buildings	Issues	Images: Ther	AlexNet	92.1% A
[124]	Buildings	Issues	Images: RGB	Fire-Det Based On EfficientNetV2, Spp-net and YOLOV3	78.1 P 76.1 mAP
[125]	Buildings	Issues	Images: RGB	DexiNed and RCF	61.50% F1-m
[126]	Buildings	Issues	Images: RGB	Faster R-CNN based ZF-Net	78.2% A
[127]	Buildings	Issues	Images: RGB	CycleGAN	NA
[128]	Buildings	Issues	Images: RGB	Deblur-DetNet	0.553 mAP
[129]	Buildings	Issues	Images: RGB	YOLOV8-based	96.2% mAP
[130]	Buildings	Issues	Images: RGB	YOLOV5x	93.2% mAP@0.5
[131]	Buildings	Issues	Images: RGB and 3D models	YOLOV5 Detection	85% F1-m 81% A
[132]	Buildings	Issues	3D models	Deng entropy combined with a bLSTM	98% A
[133]	Buildings	Issues	Predictors	MLRM	Less 3%-4% relative errors
[134]	Buildings	Issues	Predictors	UMR	0.73 P 0.82 R 0.53 NMI
[135]	Buildings	Issues	Predictors	TFT	NA
[65]	Buildings	Issues	Predictors	Context Dimension Tree	NA
[136]	Buildings	Issues	Predictors	Analytic Network Process with Zero-One Goal Programming	NA
[137]	Buildings	Issues	Predictors	NA	NA
[138]	Buildings	Issues	Predictors	Autoencoder and KNN	90% A
[139]	Buildings	Issues	Predictors	Edge-cloud architecture, smart objects with IoT sensors, and machine learning	NA
[140]	Buildings	Issues	Predictors	M5P algorithm	0.9958 R2 5.98% RAE
[141]	Buildings	Issues	Images	MobileNetV3Small	63.46% A
[142]	Buildings	Issues	Predictors	LSTM	0.078 MAE 0.005 MSE
[143]	Museums	Visitors	Predictors	LSTM Autoencoder	75.65% SA 86.89% PA
[144]	Museums	Visitors	Predictors	Non-negative matrix factorization	1.1 and 2.4 RMSE

3.4 Discussion

The previous considerations show that the use of AI and ML for CH protection is highly promising, yet several limitations and open issues remain.

One of the main issues is the lack of public datasets in the CH domain [145]. Published approaches are often developed and tested on custom datasets that are

rarely released, partly because cultural institutions are often reluctant to share their data [146]. This results in a lack of public, standardised benchmarks, making it difficult to obtain suitable training data and to compare methods fairly. In this context, the creation and release of curated datasets would already represent a substantial contribution.

This observation leads to a second point. Most of the approaches reported in the literature are DL-based. CNN architectures such as U-Net, Mask R-CNN, and ResNet are frequently adopted. These methods are often supervised and require large, carefully annotated datasets, which are often unavailable in this domain for the reasons discussed above. Moreover, sustainable AI is rarely considered in CH, despite increasing attention to this topic in other fields [147] [148]. Sustainability is particularly relevant for CH because reducing energy consumption and promoting ethical data practices directly support the long-term protection of heritage assets for future generations [149]. In addition, sustainability-oriented approaches often require shorter training times and fewer computational resources, making them accessible across different levels of resource availability while enabling strong performance with a lower environmental impact.

The growing adoption of 3D modelling, UAVs, and LiDAR data for CH monitoring can also be observed alongside limited integration across tools and disciplines. Digital heritage is being developed through multiple, often parallel, streams: IoT, which has long been integrated into monitoring systems; orthophotos and floor plans, widely used for analysing structures through RGB, thermal, and other imagery acquired by UAVs and cobots; LiDAR laser scanning and photogrammetry, which support the creation of BIM and HBIM models; and AI systems that can recognise characteristics and problems in structures and artefacts. Researchers, however, often focus on individual components, whereas the development of interoperable platforms is less common. Integration efforts are becoming more frequent, especially in the context of Digital Twins, but they are still often conceptual. One example is the limited integration of simulation models and tools for generating what-if scenarios, which are crucial for strategic forecasting in the maintenance of CH assets.

In this context, generative AI is gaining importance for anomaly detection in CH protection, but its use remains largely confined to these specific goals.

Since LLMs have shown limitations, including errors caused by insufficient semantic understanding, recent work has explored integrating LLMs with ontologies, which make domain knowledge explicit and can improve system reliability [149]. Such ontologies can also be incorporated into system knowledge bases to enable efficient access to structured knowledge and to support interoperability with contextual and CH-related data.

It is also worth noting that most existing work focuses on using AI for anomaly detection, the automatic identification of maintenance issues, and vis-

itor support. However, relatively few studies investigate how to integrate these technologies into comprehensive information and decision-support systems for CH protection that actively assist users in decision-making and situation recognition [150]. Digital Twin-based approaches move in this direction, but they rarely account for the human component and the way users interact with the physical world being monitored.

Indeed, the integration of SA paradigms [20] [151] into CH decision-support systems is recent and remains largely unexplored, except for a small number of studies that focus primarily on enabling technologies rather than on embedding SA principles into operational systems [139]. A promising direction is the development of systems that genuinely support operators' situation awareness, while remaining generalisable across different CH settings, including museums, libraries, archaeological sites, and other contexts. Such an approach should cover the full decision-support pipeline, from data acquisition and analysis to comprehension and projection, the definition of a knowledge base, and the development of applications and user-centered interfaces that support maintenance activities.

3.5 Summary

This chapter reviews the state-of-the-art literature on AI applied to CH protection, examining how AI-based techniques are currently used to enhance the safeguarding of CH assets.

Our methodological approach comprises four main steps: (i) *search*, using engines and databases such as Scopus, IEEEXplore, ScienceDirect, and SpringerLink with a targeted keyword query; (ii) *selection*, retaining only relevant journal articles and conference proceedings published in English since 2015, focused on computer science, engineering, and mathematics, and directly related to CH safeguarding and protection; (iii) *trend analysis*, investigating countries and institutions of origin, the most frequent keywords, and the most active journals, supported by a bibliometric analysis performed with VOSviewer; and (iv) *technical analysis*, defining taxonomies that describe the classes of entities studied, the problems addressed, and the techniques applied for each purpose, while highlighting methodological limitations and situating the overall research landscape.

Cultural heritage can be grouped into two macro-categories, namely *movable* and *immovable* heritage. For movable heritage, the main objectives include damage detection, entity identification, and reconstruction or restoration. Typical data sources comprise images acquired with different modalities, including RGB, X-ray, infrared, and hyperspectral, as well as predictors derived from

sensor measurements. Many contributions rely on pretrained networks such as YOLO, Mask R-CNN, U-Net, and ResNet, on customised convolutional architectures, on traditional machine-learning techniques, and on GAN-based solutions for restoration. For immovable heritage, commonly used data include imagery, laser scanning products, and sensor readings. The most frequent objectives are entity identification, issue detection, and visitor support. Widely adopted methods include Mask R-CNN, U-Net, Faster R-CNN, and YOLO-based models, which are mainly employed for detection and segmentation tasks.

The discussion highlights several persistent limitations and open questions. A central issue is the scarcity of public, standardised CH datasets. This is partly due to institutional reluctance to share data, which limits reproducibility and hinders fair benchmarking. Moreover, most deep-learning methods in the field are supervised and require large annotated datasets, which are often unavailable in CH applications. Sustainable AI considerations are still limited in CH, despite the need to reduce energy consumption and to support ethically grounded practices. In addition, integration among IoT, UAV, LiDAR, BIM/HBIM, and AI components remains insufficient. Many efforts focus on standalone tools rather than interoperable platforms. For instance, Digital Twins are often presented at a conceptual level and only rarely connected to simulation engines that enable what-if analyses.

The adoption of generative AI is currently mostly confined to anomaly detection, while large language models still exhibit limitations in semantic understanding. Nonetheless, coupling LLMs with ontologies may improve reliability. Finally, although a substantial body of work addresses anomaly detection, automated diagnosis of maintenance issues, and visitor support, relatively few studies integrate these technologies into comprehensive decision-support systems that actively assist users in decision-making and situation recognition. The use of situation awareness paradigms in such systems is recent and remains sparsely explored. In particular, ensuring continuous operator awareness in large and complex environments, such as archaeological sites, is still largely unaddressed.

Chapter 4

Situation-aware Cyber-Physical-Social System for Cultural Heritage

"Everyone designs who devises courses of action aimed at changing existing situations into preferred ones."

— Herbert A. Simon, *The Sciences of the Artificial*

The purpose of the thesis is to support the decision-making of human operators in the field of CH preservation using new technologies and emerging approaches. These maintenance processes, as seen in our reference lifecycle in Section 2.1.1.1, are usually guided by surveys of the CHCI, traditionally performed by humans using traditional methods. Nowadays, such surveys can be supported by aerial and ground drones in order to identify maintenance issues [152]. Such a process is human–machine since the human component is present and necessary. Also the decision-making processes regarding maintenance planning are guided by humans, which, in order to decide, take into account the vulnerabilities of the various structures or works of art within the CHCI, by calculating vulnerability indices, like those in equations 2.1 and 2.2. However, updating such indices in a timely manner remains challenging. Even when surveys are conducted using UAVs, they are often carried out at long intervals. For example, at the archaeological site of Pompeii, the aerial surveys are conducted monthly, leaving operators blind until the next flight. The process of identifying maintenance issues based on these photos is also long and complex. These procedures can be made more systematic and machine assisted. The identification of maintenance issues can be automated using AI-

based approaches, and even when long intervals separate successive surveys, combining information from the most recent imagery with real-time environmental data, such as weather and environmental conditions, allows operators to remain aware of ongoing situations and to understand how existing issues are evolving or may evolve [27].

Considering the interaction among the physical CH asset; the devices and software used for its analysis and monitoring; and the human operators, domain experts, and visitors of the asset, our proposal is to view the CH asset as a Cyber–Physical–Social System [153] [154]. In this way, human operators are supported in planning their maintenance activities, as they are provided with an up-to-date global picture of the situation.

We propose a Situation-aware CPSS methodological approach for cultural heritage protection. This SA-based approach enables effective operator support even over the long timescales typical of cultural heritage monitoring by combining information collected during surveys with real-time contextual data, thereby maintaining operators' awareness of ongoing conditions and vulnerability of the zones. To the best of our knowledge, this is the first time that an SA-based approach has been proposed to support the maintenance of CHCIs.

Parts of this chapter have been previously published in:

- Francesco Colace, Giuseppe D'Aniello, Massimo De Santo, Rosario Gaeta, Gabriel Zuchtriegel. *Situation-aware Cyber–Physical–Social System for Cultural Heritage*. Intelligent Systems with Applications. 2025. Volume 27. https://doi.org/10.1016/j.iswa.2025.200544.
- Luca Aliberti, Giuseppe D'Aniello, Massimo De Santo and Rosario Gaeta. *A Situation-Aware Cloud-Edge Architecture for Cultural Heritage Management*. IEEE COGSIMA. 2025. doi:10.1109/CogSIMA64436.2025.11079541.
- Francesco Colace, Massimo De Santo, Rosario Gaeta, Rocco Loffredo, Luigi Petti. *FAUNO: A Machine Learning-Based Methodology for Monitoring and Predictive Maintenance of Structures in Archaeological Parks Through Image Analysis*. HCII. 2024. https://doi.org/10.1007/978-3-031-60611-3_24.

4.1 Related Works

From the literature, we note that the application of SA in the CH domain is in its very early stages. In our work, we propose a novel approach to support SA for the maintenance of CH assets. Before discussing our methodology, we analysed similar applications of CPSS for maintenance activities and approaches in the CH world.

4.1.1 Cultural Heritage Maintenance Approach

Fan et al. [155] introduced the M-RM system, a metaverse-enabled road maintenance system based on CPSSs, implementing the simulation, monitoring, diagnosis, and prediction functions of road systems in the metaverse using human and computer information. The AIDA algorithm is used for data augmentation in combination with a crack detection approach named PAT, which combines frequency-division convolution and a mixed-domain attention mechanism. Meng et al. [156] introduced a framework to support data mining in CPSSs, namely Multidimensional time series Outlier detection based on a Time Convolutional Network AutoEncoder (MOTCN-AE), which can detect outliers in time-series data to identify equipment failures, dangerous behavior of cars, and other issues. The approach in this paper transforms time series into a feature-rich time series, and then uses the original TCN-AE approach to reconstruct the feature-rich time-series data; the reconstruction error is used to calculate outlier scores. The framework has been validated with various time-series datasets. The work of Faramondi et al. [157] proposed a novel unsupervised cyber–physical anomaly detection framework which implements a combination of various anomaly detectors using a Bayesian-network-based probabilistic model. The Bayesian network is used to assess the trustworthiness of the anomalies detected through two anomaly-detection modules capable of monitoring, at the same time, the behavior of cyber and physical data acquired. De Benedictis et al. [10] introduced an architecture for Industrial IoT anomaly detection based on the paradigms of digital twins and autonomic computing (AC). The approach is based on the MAPE-K feedback loop of AC in order to monitor, analyse, plan, and execute appropriate reconfiguration or mitigation strategies when a failure happens or may happen.

4.1.2 SA for Cultural Heritage

In Nota et al. [139], the authors have recently proposed an SA model for heritage building conservation and a cyber-physical system to support conservation processes. As a case study, the authors proposed an implementation of such a cyber-physical system for monitoring the Royal Palace of Carditello. Casillo et al. [138] combine autoencoders and IoT technology for monitoring, predictive maintenance, and decision-making to protect CH buildings. The autoencoder is used to detect damage, while k-NN classifies it. Restoration works are determined by the Decision Support module. Rodrigues et al. [158] use a Region-Based Convolutional Neural Network to perform automated image-based surveys.

The authors focused on automatizing the recognition and the classification of defects in historical buildings, to finally interpolate this geometric and numerical information with their BIM. Cecere et al. [159] realized a digital twin through the integration of data coming from the various sensors in the HBIM model, exploiting real-time IoT devices placed on the structures to be monitored. The proposed methodology was applied to a real case study within the Velia Archaeological Site: Porta Rosa, which is the oldest known example of a round arch in Italy. In Qiu et al. [98], to preserve historical buildings in southern Fujian, China, a method based on an improved YOLOv8 [160] neural network is proposed to identify roof damage from aerial imagery in the cities of Xiamen and Quanzhou, Fujian Province. Based on high-resolution orthophotographs obtained from unmanned aerial vehicle (UAV) tilt photography, the YOLOv8 model was used to make predictions. The approach first uses a YOLOv8-based architecture to identify buildings in the region. Once the structure has been identified, another YOLOv8-based model is used to identify damage on the roofs.

4.2 Situation-aware CPSS

In Figure 4.1, our novel Situation-Aware Cyber-Physical-Social System methodological approach for general cultural heritage sites is illustrated. The Physical-Layer is composed of the CH site to be maintained and the actuators of the maintenance process. The Cyber-Layer is made up of devices for data acquisition and analysis to support the SA and decision-making of humans, who compose the Social-Layer of the system, consisting of experts, operators, and visitors. The experts are humans who are in charge of studying artifacts, structures, materials, and construction techniques, and who are capable of evaluating the vulnerability of the asset. Mathematical metrics and indices are

used to measure the vulnerability of artifacts and structures and to determine priorities. The human operators, instead, are in charge of maintaining the asset and addressing issues.

In the following, we will detail the various layers of the CPSS, with particular attention to the Cyber-Layer.

4.2.1 Physical-Layer

The Physical-Layer is precisely the level we aim to protect, and it is the direct focus of maintenance interventions. It represents the tangible cultural heritage site and the environmental context in which it is embedded. It includes built structures such as walls, columns, pavements, stairways, protective roofs, and shelters, together with in situ and movable artefacts and all material components whose integrity is essential for conservation and public safety. Unlike controlled indoor heritage settings, a cultural heritage site is typically exposed to open-air conditions and a complex morphology, which can amplify degradation processes and emergency scenarios.

From a risk perspective, the Physical-Layer is continuously influenced by slow and sudden stressors. Long-term deterioration is driven by weathering and microclimate variability, including thermal cycles, wind, precipitation, humidity, salt crystallization and biological colonization. Acute events can rapidly push the site into unsafe conditions. Examples include heavy rainfall leading to water accumulation and erosion, strong winds affecting temporary protections, heatwaves increasing thermal stress and visitor vulnerability, and natural hazards such as floods, landslides, wildfires, and earthquakes. These conditions may cause detachments, cracking, instability of structural elements, obstruction of pathways, and reduced accessibility, which require timely detection and response.

To capture these dynamics, the Physical-Layer incorporates fixed and mobile sensing integrated within an IoT infrastructure. Fixed sensing nodes can monitor microclimate, structural and geotechnical indicators, and safety-related signals. Mobile sensing platforms, including aerial or ground cobots and handheld devices used by staff, support close-range inspections, rapid post-event surveys, and coverage of areas that are difficult to access safely.

Beyond sensing, the Physical-Layer may include actuators that enable timely mitigation actions directly on-site. Typical examples include automatic or remotely controlled barriers and gates to prevent access to unsafe or sensitive areas, dynamic signage and beacons to guide visitor flows, and warning devices such as lights or sirens. Cobots can also support targeted interventions under human supervision, for example by assisting in debris removal after

minor incidents, handling tools for rapid temporary securing, and performing repetitive low-risk inspection and maintenance tasks.

The resulting sensor streams, contextual information, and actuation capabilities provide the primary basis for SA, predictive assessment, and the activation of mitigation measures.

4.2.2 Social-Layer

The Social-Layer captures the human actors interacting with the CH site and the digital mechanisms that coordinate their actions. These actors form a social network: they continuously interact, exchange information, and cooperate across roles (e.g., analysts, conservators, maintenance operators, and managers). This networked behavior is not incidental but operationally necessary, because observations collected on-site must be interpreted, validated, translated into decisions, and converted into interventions. Similarly, visitors can be informed about what is happening and the current situation so that they can adjust their movements, and they themselves can guide the situation perceived by the cultural site administrators with their route. The Cyber-Layer, which will be described later, plays a key role in this cooperation by enabling traceable communication and coordinated workflows across actors.

The Social-Layer includes the stakeholders who interpret information, make decisions, execute interventions, and experience the site as visitors. A first group consists of domain experts such as archaeologists, conservators and restorers, heritage scientists, and site managers who perform inspections and validate the meaning of sensor observations. Their role is essential to connect measured signals such as cracks, vibrations, and moisture patterns to conservation-relevant interpretations, and to define priorities, thresholds, and acceptable risk levels. Complementary to them are maintenance and operational staff. This group includes field operators responsible for practical activities such as minor repairs, cleaning, securing areas, placing signage, and checking sensing equipment. It also includes planning and supervisory roles responsible for scheduling interventions, allocating resources, coordinating contractors, and ensuring compliance with safety and conservation protocols. This division supports rapid on-site response as well as consistent long-term maintenance planning.

Visitors represent a third crucial actor category. They are potential targets of risk in the presence of falling fragments, slippery surfaces, heat stress, and crowding during evacuations. They may also be a source of impact on heritage, either unintentionally or intentionally. At the same time, visitors can be leveraged as *moving sensors*. Aggregated information on flows, densities,

and spatial patterns can highlight anomalies such as access to restricted areas, unusual clustering near fragile structures, or abrupt changes in movement that may indicate incidents. For this reason, the Social-Layer must support two-way interaction. It should inform visitors through clear guidance and real-time alerts when hazardous conditions emerge, while enabling site staff to use visitor-derived indicators to enhance SA and trigger timely protective actions.

4.2.3 Cyber-Layer

The Cyber-Layer provides the digital backbone that connects the physical world of the cultural heritage site with the involved human actors. It aggregates heterogeneous data streams from fixed and mobile sensors, contextual information, and human inputs, and transforms them into actionable knowledge to support monitoring, decision-making, and timely interventions. In this way, the Cyber-Layer enables the transition from raw observations to higher-level situation awareness, supporting anomaly detection, the assessment of evolving conditions, and the activation of safety and preservation measures.

Moreover, the Cyber-Layer acts as an integration and orchestration space. It hosts data management services, models, and interfaces that coordinate stakeholders and operational processes. It supports both real-time and offline analysis, integrates baseline domain knowledge, and provides mechanisms for communication and alerting for site staff and visitors.

To support human decision-making for such an asset, we propose a Cyber-Layer structured around the three levels of situation awareness, (L1) perception, (L2) comprehension, and (L3) projection, according to Endsley's model of SA [20].

The details of the framework proposed for the Cyber-Layer are identified in Figure 4.1. The whole process must be intended as iterative:

- **L0 Sensing:** data acquisition, cleaning and storage coming from sensors (such sensors can be both physical or virtual). The data will be consequently used to feed the perception phase. The phase can be divided into three sub-phases:
  - **L0.1 Acquisition:** the acquisition of sensor data. This acquisition can be carried out with the support of ground or aerial cobots;
  - **L0.2 Offline:** considers manual or semiautomatic preparation of the data (like data cleaning, merging, georeferencing and storage) that, in some context, depending on the system configuration and the type of sensors the system has, must be done offline;
  - **L0.3 Online:** data cleaning and storage operation that can be done

in a real-time manner.

- **L1 Perception:** data processing to identify damages, issues, maintenance needs, and context states. This level consists of three modules:
  - **L1.1 Low-level perception:** data pre-processing, segmentation and feature extraction;
  - **L1.2 High-level perception:** dimensionality reduction and feature selection;
- **L2 Comprehension:** this level performs situation detection on the result of relief, using the features extracted and selected in the perception phase, and then fuses such information with the perceived context. In particular:
  - **L2.1 Situation Detection:** this phase applies machine learning models to detect position and type of maintenance issues. If new relief are not available, such module uses default knowledge situation, stored in the previous iterations. The maintenance issue can later be combined with contextual info to understand the evolution of the situation;
- **L3 Projection:** tracking the possible evolution of the situations via predictive models, updating vulnerability indices.
  - **L3.1 Situation Projection:** this phase is concerned with understanding how vulnerability and risk on various areas of the CH asset will change based on perceived and understood information. Contextual data such as time, location, environmental condition are used together with the situation identified to calculate the vulnerability indices using the Context Space Theory (CST);
  - **L3.2 Danger Projection:** by also considering projections of the system's evolution, rule based approaches, as well as machine learning, can be used to anticipate potential criticalities and hazards that may affect the future condition of the CHCI. This enables the system to provide recommendations to human operators, supporting their decision-making. Based on what is perceived and on the current goal, the system can suggest appropriate actions to the user or redirect the current goal when necessary.

Each phase of our model, which is illustrated in Figure 4.1, is described in the following sections:

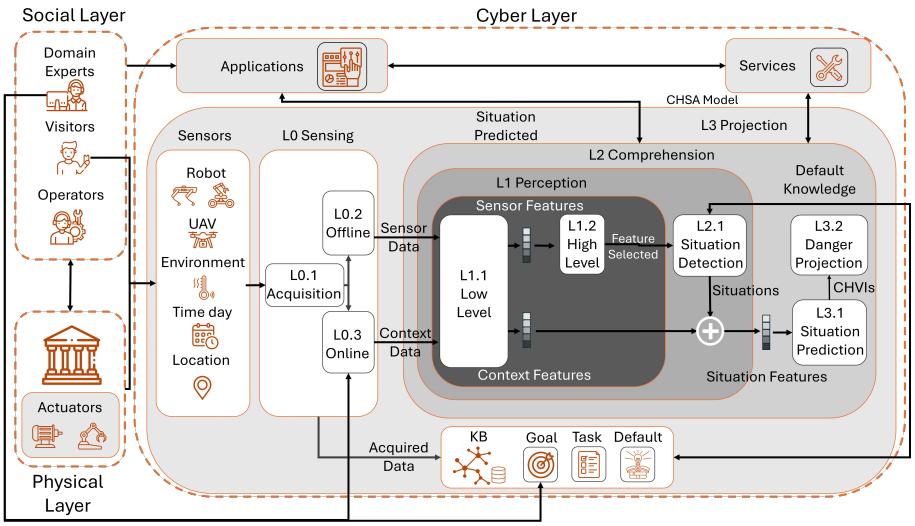


Figure 4.1: Situation-aware Cyber-Physical-Social System for Cultural Heritage Protection

4.2.3.1 L0 Sensing

The sensing phase manages the data acquisition process, which involves data cleaning and data storage activities.

- **L0.1 Acquisition:** acquiring heteregeneous data from various types of sensors (e.g. cameras, environmental sensors) at different time frequencies, can be challenging. In the CPSS a set $S = s_1, s_2, ..., s_n$ of n sensors is considered. A data stream Δ_i is acquired from each sensor $s_i \in S$, with $i = 1...n$. Each data stream

$$\Delta_i = [\delta_i^0, \delta_i^1, ..., \delta_i^l, ...] \tag{4.1}$$

with δ_i^l gathered by the sensor s_i at the time instant $t = l$ with $l \in \mathbb{R}$. In particular, δ_i^l is a general m-dimensional matrix.

$$\delta_{\mathbf{i}}^{\mathbf{l}} = \begin{bmatrix} \begin{bmatrix} \beta^l_{i,1,1,1}, \beta^l_{i,2,1,1}, \ldots, \beta^l_{i,k,1,1} \\ \beta^l_{i,1,2,1}, \beta^l_{i,2,2,1}, \ldots, \beta^l_{i,k,2,1} \\ \vdots \\ \beta^l_{i,1,h,1}, \beta^l_{i,2,h,1}, \ldots, \beta^l_{i,k,h,1} \end{bmatrix}, \\ \vdots \\ \begin{bmatrix} \beta^l_{i,1,1,m}, \beta^l_{i,2,1,m}, \ldots, \beta^l_{i,k,1,m} \\ \beta^l_{i,1,2,m}, \beta^l_{i,2,2,m}, \ldots, \beta^l_{i,k,2,m} \\ \vdots \\ \beta^l_{i,1,h,m}, \beta^l_{i,2,h,m}, \ldots, \beta^l_{i,k,h,m} \end{bmatrix} \end{bmatrix} \quad (4.2)$$

where $k, h, m \in \mathbb{N}$ and $k, h, m \geq 1$ and the value of k, h, m depends on the type of sensors (if it is a sensor that acquires a single value in one instant, the matrix corresponds to a single value, while for other types of sensors, like cameras, the matrix will have dimensions $\neq 0$). Then, defining the data stream $\Delta = \{\Delta_1, \Delta_2, ..., \Delta_n\}$ it can be divided into two types of data streams that deal with two different types of preprocessing activities: Δ^s and Δ^c where $\Delta^s \cap \Delta^c = \varnothing$. Such operations can be done offline or online.

- **L0.2 Offline:** in the CH domain, there are often experts who periodically conduct surveys using aerial and ground cobots equipped with cameras or LiDAR sensors. The outputs of such devices must be processed using specific tools to obtain georeferenced orthophotos or 3D models. Nowadays, this work has to be done by humans. The set of data going through such a manual process is Δ^c. Each data stream in Δ^c undergoes a transformation process guided by a human. Indeed, the SA system can also support these humans in their operations; for example, the system can suggest how to resolve occlusions or alignment issues;
- **L0.3 Online:** operations like data cleaning, data imputation, redundancy removal, and outlier removal reduce possible problems in the following phases. Such operations involve the stream Δ^s, which must be analysed as a time series and does not need human intervention in its pre-processing.

4.2.3.2 L1 Perception

The perception is the first phase of the Endsley SA model and consists of the data processing after the sensing phase. In particular the subphases are explained in the following paragraphs:

- **L1.1 Low-level Perception:** the main task of the low-level perception

phase involves extracting features from sensor data, intended for issues and anomaly detection, but also for contextual data used to identify the context states. Considering the stream $\Delta^s = \{\Delta_1^s, \Delta_2^s, ..., \Delta_o^s\}$ and $o \leq n$ of contextual data sensors fusing and processing such data streams, we can extract a series of features $\Phi^s = [\phi_1^s, \phi_2^s, ..., \phi_m^s]$ with $m \leq o$. Often, each data stream Δ_i^s must be divided into time windows:

$$W_{\Delta_i^s}^j[\cdot] = \Delta_i^s[j \cdot size, ((j+1) \cdot size) - 1] = \delta_i^{j \cdot size}, \delta_i^{((j \cdot size)+1}, \delta_i^{((j+1) \cdot size)-1} \quad (4.3)$$

where $W_{\Delta_i^s}^j[\cdot]$ is the j^{th} window which contains the data points extracted from Δ_i^s in the interval $[j \cdot size, ((j+1) \cdot size) - 1]$. The feature value $\phi_l^s(j), l = 1, ..., m$ represents the value of the l^{th} feature. It is computed by applying a transformation function $f_{\phi_l^s}$ on the window $W_{\Delta_i^s}^j$

$$\phi_l^s(j) = f_{\phi_l^s(W_{\Delta_i^s}^j)} \quad (4.4)$$

The features can also be computed by aggregating and processing data from multiple windows across different sensors. Afterwards, such features go in input to the L1.3 Context recognition module.

Looking at the stream $\Delta^c = \{\Delta_1^c, \Delta_2^c, ..., \Delta_q^c\}$ and $q \leq n$, various traditional or deep learning techniques can be applied on the result of the surveys to obtain the set of features $\Phi^c = [\phi_1^c, \phi_2^c, ..., \phi_u^c]$ with $u \leq q$. This set goes in input to the L1.2 High-level perception phase.

- **L1.2 High-level Perception:** at this point, a feature selection is performed on the set Φ^c of u features. This step helps reduce the complexity of the problem, both for computational complexity and performance reasons. In particular, a subset $F^g \subseteq \Phi^c$ of g features is produced. To select such features various approaches, including statistical and learning-based methods, can be used.

4.2.3.3 L2 Comprehension

The primary objective of level L2 is to conduct situation identification. The data acquired and appropriately fused from the perception phase are used to derive situations and to support the user goals [161]. In particular, we define two sub-phases:

- **L2.1 Situation Detection:** this phase uses the g features F^g to identify the presence of situations of interest, including defects, collapses, hazards, and other anomalies, between r of classes $c \in C = \{c_1, c_2, ..., c_r\}$,

coincident with the set of goals identified by the domain experts. In particular, let $Z = \{z_1, z_2, ..., z_y\}$ be the set of all the zones of a CH property, that can be, for example, areas or structures in an immovable property or section in a movable property, with $j = 1, 2, ..., y$, the situation detection phase uses an approach that for each z_j outputs $M_j = \{m_{j,1}, m_{j,2}, ..., m_{j,u}\}$ with u number of maintenance issues found, that is the set of all the issues found in that zone z_j: in particular, given $k = 1, 2, ..., u$ each $m_{j,k} = [position, c_r]$ is characterized by the position and class of the problem.

4.2.3.4 L3 Projection

Once the situations in the CH entity have been detected and such information has been integrated with contextual info, the projection phase must focus on how such contextual info can be used to understand the evolution of the situation and what parts of the CH entity are more vulnerable [162].

- **L3.1 Situation Projection:** the vulnerability evolution of each zone of the cultural heritage site is assessed by leveraging the CST together with the identified maintenance issues. Zone vulnerability is quantified through specific indices, whose computation will be detailed in section 4.3. By combining automatically detected maintenance issues with environmental and social contextual information, these indices enable the system to update the state of the cultural heritage asset over time, even in the absence of recent surveys for a given zone;

- **L3.2 Danger Projection:** the update of the vulnerability status of the structures allows the human operators to understand which structures should be prioritized for inspections and maintenance. However, the vulnerability index may not be sufficient to enable rapid intervention when the situation requires it. Contextual information should be used to feed a predictive model that can issue timely alerts, prompting urgent surveys and maintenance in specific zones. For example, if the predictive model detects an increased likelihood of a severe weather event during a given period, it could warn operators to secure areas that may be heavily damaged. Indeed, even if such an issue would already be taken into account in the vulnerability index, this urgency should be signaled in advance, not least because vulnerability to the adverse event could still be lost within the calculation of that index. In this way, we are able to encourage the user to change their goal if necessary.

4.2.3.5 CPSS mapping to CH Lifecycle

The SA CPSS for CH can support all the steps necessary for the maintenance activity of a CH asset, as in Figure 2.1. The *step S1: Preliminar survey* is a phase which is not inside the maintenance lifecycle of the CH asset. However, the result of such phase, together with the *step S2: Detailed Analysis* act as first level of perception and background knowledge useful in calculating the vulnerability index of the zones in the site. The *step S3: Risk assessment* is performed by the high level perception and situation detection modules of our CPSS. The contextual information and the risks identified are then processed and put together in the *step S4: Intervention Planning* where the updated vulnerability risks are used to understand prioritization of intervention and how the intervention must be performed, which is fundamental in the Projection modules. The applications and services of the CPSS must then be used to perform the *step S5: Intervention execution* Cyclically the *step S6: Monitoring Survey* is executed by the Sensing and low-level perception modules, and starts a new monitoring cycle.

4.3 Cultural Heritage Vulnerability Index and Situation Modeling

Any SA system requires a computational model of the situation, including the SA-CPSS introduced in this work. Cultural heritage protection and maintenance are governed by well-established standards and regulations. As discussed in Section 2.1.1.2, both the literature and current archaeological practice rely on vulnerability measurements of different site areas to prioritize maintenance interventions. Experts define and compute these indices by directly assessing site conditions and determining the relative contributions of area characteristics. In this domain, vulnerability assessment therefore plays a central role in defining intervention priorities and in helping experts understand the conditions of individual zones as well as the site as a whole.

Equations 2.1 and 2.2 present recent computational models for assessing the vulnerability of archaeological site areas. However, these models are static and depend on periodic expert surveys, during which contribution values are assigned based on observed conditions, site characteristics, and materials. Consequently, operators remain unaware of ongoing site conditions until the next survey is performed.

In contrast, we propose a situation model in which area vulnerability is not assessed solely through expert evaluations conducted during surveys, but is continuously updated by integrating automatically detected maintenance issues with relevant contextual information. This approach provides an up-to-date

representation of site conditions during the intervals between surveys.

In this regard, to model such situations, for each zone z_j in a cultural heritage site we use eq. (2.2) to define $CHVI_j$ as its vulnerability index associated. Given this, we consider as vulnerability index:

$$CHVI_j = w^s \cdot CHVI_j^s + w^d \cdot CHVI_j^d \tag{4.5}$$

composed of a static and a dynamic component weighted by w^s and w^d: the static component relies on the vulnerability index proposed by [34] which sees $CHVI^s$ as:

$$CHVI_j^s = E_j + S_j - R_j \tag{4.6}$$

where E_j is the exposure, S_j is the susceptibility, and R_j is resilience for the zone z_j. In particular E_j can be calculated from a series of n_e criteria (e.g. Cultural Significance, Economic conditions, Infrastructures etc.) cr_e, each of this characterized by a series of m_e subcriteria scr_e (e.g. Cultural traditions, Cultural acknowledgments etc.). In particular:

$$E_j = \sum_{i=1}^{n_e} \gamma_{cr_e,i} \left(\sum_{k=1}^{m_e} (value_{e,k}) \gamma_{scr_e,k} \right) \tag{4.7}$$

S_j can be calculated from a series of n_s criteria cr_s (e.g. Buildings, Vegetation etc.), each of this characterized by a series of m_s subcriteria scr_s (e.g. Materials, Use, State of Conservation etc.). In particular:

$$S_j = \sum_{i=1}^{n_s} \gamma_{cr_s,i} \left(\sum_{k=1}^{m_s} (value_{s,k}) \gamma_{scr_s,k} \right) \tag{4.8}$$

R_j can be calculated from a series of n_r criteria cr_r (e.g. Preparedness, Coping, Restorative capacity etc.), each of this characterized by a series of m_r subcriteria scr_r (e.g. Maintenance, Policy regulation, Physical recovery etc.). In particular:

$$R_j = \sum_{i=1}^{n_r} \gamma_{cr_r,i} \left(\sum_{k=1}^{m_r} (value_{r,k}) \gamma_{scr_r,k} \right) \tag{4.9}$$

where each $\gamma_{cr_r,i}$ is the coefficient for cr_i, each $\gamma_{scr,k}$ is the coefficient for the subcriteria scr_k, and $value_k$ is the value given to the subcriteria scr_k.

Each of these components is calculated by experts based on a series of criteria and sub-criteria, whose values and weights are defined by the experts themselves, who, based on their experience, are able to assess, measure, and compare the condition of the various areas of the site.

In order to gain a thorough and updated understanding of the vulnerability of certain structures, our proposal is to take into consideration the maintenance

issues at the site together with contextual information, using AI and Context Space Theory (CST).

CST is a specification-based technique proposed by Padovitz, Boytsov et al. [163] that employs a geometric metaphor to provide a clear and insightful representation of context. In CST, context is viewed as a multidimensional space, with each axis, denoted x_i, representing a key feature of the context. The value on a context-space axis represents a context attribute value, and a complete set of pertinent context attribute values at a specific time constitutes the context state X. This context state defines a point within the context space. Figure 4.2 shows an example of a three-dimensional space defined by temperature (t), humidity (u), and precipitation (p) as context attributes.

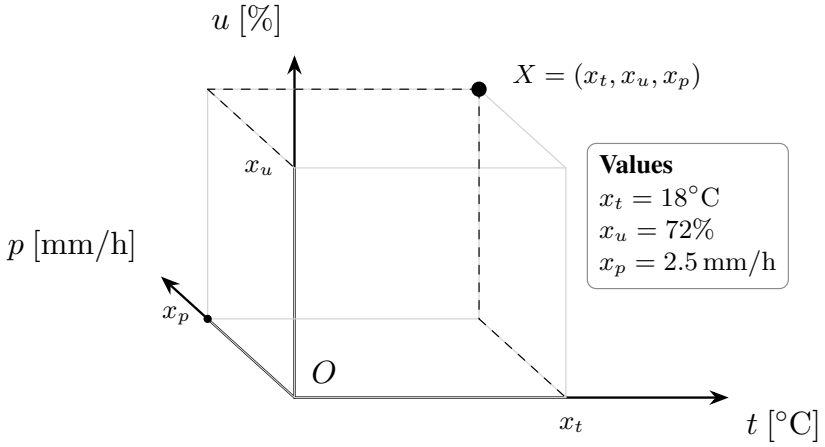


Figure 4.2: Example of 3D contextual space.

In our case, the context state is not represented solely by environmental data but also includes the number of detected maintenance issues and other contextual factors, such as the number of visitors. The combination of all contributions related to these context attributes enables the dynamic update of a structure's vulnerability. The $CHVI^d$, and so the whole $CHVI$ of a structure is used to calculate the predicted vulnerability of such structure. So, the dynamic component of the vulnerability index $CHVI^d$ can be calculated as:

$$CHVI_j^d(X) = \sum_{i=1}^{N} w_{i,j} \cdot con_j(x_i); \tag{4.10}$$

$$con_j(x_i) = \begin{cases} v_{j,1}, & x_i \in I_{i,j,1} \\ v_{j,2}, & x_i \in I_{i,j,2} \\ & \dots \\ v_{j,m}, & x_i \in I_{i,j,m} \end{cases} \tag{4.11}$$

where:

- $CHVI_j^d$ represents the risk level for z_j in the context state X;
- The function $con_j(x_i)$ the contribution of the i^{th} context attribute value to the risk of the zone z_j. Usually, this function is a step function over a certain context attribute, and in this case, $I_{i,j,k}$ are the generic intervals related to the h^{th} context attribute. Each interval is associated with a contribution value $v_{j,k}$ to the situation, with $v_{j,k} \in [0, 1]$.
- $w_{i,j}$ represents the relative weight of the i^{th} context attribute for z_j. Such weight depends also on the maintenance issues already founded in that part of the CH entity.

The definition of these contribution functions is critical because it determines how strongly each attribute affects the vulnerability of an area and, consequently, the operator's decisions. For this reason, operators often want direct control over these functions. Even if the index computation becomes automatic once the functions are defined, they remain responsible for determining what constitutes a risk situation. In our case, the contribution functions and the weights associated with each attribute were defined through an expert-based process, meaning they were specified by domain experts.

Nothing prevents these functions from being defined through other approaches, including data-driven methods. In such cases, the definition process may be faster, but it would reduce expert control, potentially creating resistance in sensitive contexts such as critical infrastructure maintenance.

In any case, contribution-function-based approaches offer a major advantage in terms of explainability. At any time, it is possible to trace the reasons behind a decision and understand why a risk situation in one area was considered more critical than in another, or vice versa. This property is fundamental in the context of CHCI protection.

The vulnerability index introduced in Equation 4.5 is proposed as a measure of overall vulnerability, capable of capturing the various contributions arising from all the variables that affect risk within a CHCI. In practice, however, our proposal defines a general CHVI whose weights and contribution functions can be adapted to the user's current goal. The same attribute can contribute differently to the CHVI of a zone depending on what the user is interested in at a given moment and, therefore, on the user's current goal. For such reasons, the CHVI is a goal-driven vulnerability measure. For example, rain is a highly relevant factor when assessing vulnerability related to broken tiles or damaged roofing membranes, whereas its contribution becomes minor or even irrelevant when the focus is on weedy vegetation.

In this chapter, we use a goal-independent version of the index, where each attribute contributes regardless of the user's current goal, as discussed in

Section 4.4.1. During the CPSS implementation phase, however, in Chapter 7 and Section 7.3.4, we introduce a more complete and refined goal-driven version, defining multiple goal-specific CHVIs, specifying the corresponding weights and contribution functions for each one.

4.4 Proof of Concept

We propose a proof of concept relative to the archaeological site's field to show how the model works and how it can support archaeological site operators in their decisions. The protection of archaeological sites is very critical in the CH domain, because of their size and outdoor nature. Archaeological sites can be defined as a complex of ancient structures, works of art and artifacts that need to be protected and safeguarded. The maintenance issues and defects inside an archaeological site must be detected and solved quickly to avoid severe damage to structures or their artworks, like frescoes that can be damaged by rain inside a broken cover.

As maintenance issues, we focus on roofs with broken tiles, damaged conduits and weedy vegetation. Such classes have been identified by experts of the archaeological site of Pompeii. In this proof of work, we used a YOLOV7 detector for issues identification. At this stage, however, we will not detail the approaches used to classify and detect roofs, as these will be discussed in detail in the chapter 6.3.2.

We apply our approach to three case studies that differ substantially in the characteristics of the facilities and, in our case, in the available resources:

Case Study 1. **Archaeological Site of Pompeii:** we consider four structures, namely *Casa del Fauno*, *Casa del Labirinto*, *Casa di Gavius Rufus* and *Casa di Nave Europa*, for which high-resolution aerial photos are available, acquired through periodic UAV surveys conducted at an altitude of 80 meters;

Case Study 2. **Archaeological Site of Paestum:** we considered three structures, namely *Tempio di Hera*, *Casa dei Sacerdoti*, and *Altare di Nettuno*, for which satellite imagery is available;

Case Study 3. **Archaeological Site of the Colosseum:** we considered three structures in Rome, namely *Colosseo*, *Museo del Foro Romano*, and *Basilica di San Clemente*, for which satellite imagery is available.

In Case Studies 2 and 3, the available aerial imagery is not of sufficient quality to detect broken tiles and conduits, but it does allow the detection of

weedy vegetation.

4.4.1 Cultural Heritage Vulnerability Index Calculation

We will introduce the calculation of the overview CHVI as introduced in Eq. 4.5, as a combination of the static contribution reported in Eq. 4.6 and presented in detail in section 4.4.1.1, but also a dynamic contribution, Eq. 4.10, obtained using information retrieved in the high perception phase and from the context space theory, and presented in section 4.4.1.2.

This will be the calculation used in each of the areas covered by the proof of concept proposed in this chapter.

4.4.1.1 Static CHVI Contribution

The Eq. 4.6 applies a number of criteria and subcriteria that are negligible in our case, such as the one related to economic factors: for this reason, we use only a subset of them. In particular, looking at the Eq. 4.6, for the static part we focus only on three of the susceptibility S_j criteria: for this reason, the weight linked to susceptibility is set to 1 and the others to 0. In table 4.1 such criteria are synthesized: in particular the three criteria are related to buildings, built/human-made features and vegetation. Obviously, since this vulnerability index in the reference paper [34] is between 0 and 1, and since in our work we do not consider all the sub-criteria related to vegetation, the weights linked to the vegetation sub-criteria have been slightly rebalanced.

Criterion	**Subcriterion**	γ_c	γ_{sc}
CR1.1 Buildings	CR1.1a Constructions and materials	0.40	0.50
	CR1.1b Use		0.10
	CR1.1c State of conservation		0.20
	CR1.1d Previous harming interventions		0.20
CR1.2 Built/human-made features	CR1.2a Built elements of decoration	0.25	0.20
	CR1.2b Water features		0.20
	CR1.2c Circulation features		0.20
	CR1.2d State of conservation		0.40
CR1.3 Vegetation	CR1.3a Species (Tree)	0.35	0.25
	CR1.3b Grass/shrub cover		0.15
	CR1.3c Use		0.20
	CR1.3d State of conservation		0.40

Table 4.1: Table Representing the Susceptibility Analysis Criteria and Subcriteria with Respective Weights [34].

Criterion and Subcriterion	Value Meaning	Value
CR1.1a Constructions and materials	Buildings structurally strong made of resistant materials	0.00
	Buildings structurally strong with materials prone to degradation or impact damage	0.50
	Buildings structurally weak with materials prone to degradation or impact damage	1.00
CR1.1b Use	In continuous use	0.00
	Occasional use	0.40
	Abandoned	1.00
CR1.1c State of conservation	Good	0.00
	Fair	0.18
	Poor	0.73
	Very bad	1.00
CR1.1d Previous harming interventions	Yes, previous interventions	1.00
	No interventions made	0.00
CR1.2a Built elements of decoration	Absence of elements of decoration	0.00
	Presence of elements of decoration	1.00
CR1.2b Water features	Absence of water features	0.00
	Presence of water features	1.00
CR1.2c Circulation features	Absence of circulation features	0.00
	Presence of circulation features	1.00
CR1.2d State of conservation	Good	0.00
	Fair	0.18
	Poor	0.73
	Very bad	1.00
CR1.3a Species	Species tolerant to local natural and climate threats	0.00
	Species weakly tolerant to local natural and climate threats	0.30
	Species not tolerant to local natural and climate threats	1.00
CR1.3b Grass/shrub cover	Species tolerant to local natural and climate threats	0.00
	Species weakly tolerant to local natural and climate threats	0.30
	Species not tolerant to local natural and climate threats	1.00
CR1.3c Use	Intensive land use (including urban sprawl, without natural elements)	0.00
	Intensive land use with natural elements	0.30
	Extensive land use	1.00
CR1.3d State of conservation	Good	0.00
	Fair	0.18
	Poor	0.73
	Very bad	1.00

Table 4.2: Evaluation Scale Related to The Subcriteria of Susceptibility [34].

Each of these is characterized by a subcriteria that is shown in table 4.2, with value and its meaning indicated for each subcriteria. Considering the first criteria CR1.1 regarding buildings, the subcriteria are relative to their construction and materials, use, state of conservation and the presence of previous harming, looking to the criteria CR1.2 Built/human-made features, the subcriteria are related to the presence of built elements of decoration, water features, circulation features and their state of conservation. Vegetation criteria CR1.3 has species, age, use and state of conservation as subcriteria. For this case study, we asked experts in the domain to assign to each structure the value according to precision analyses and to the history of maintenance that has taken place. For each of the affected zones, we calculated the static contribution of the vulnerability index as shown in the equation 4.8. Such contribution is static and remains the same for each of the photos of the same structures.

4.4.1.2 Dynamic CHVI Contribution

What makes the vulnerability index, and so the situation, vary over time is the dynamic contribution $CHVI^d$ (Eq. 4.10), which depends on the issues identified periodically in the pictures and combined with contextual information. In particular, in this case we have meteorological and anthropogenic flow information available in the site of Pompeii, not for the others. Considering three types of maintenance issues, damaged conduits (x_{dc}), broken tiles (x_{bt}), and weedy vegetation (x_{wv}), together with the meteorological information related to precipitation (x_p), temperature (x_t), relative humidity (x_u), and wind speed (x_w) and finally considering also the visitors flows (x_v), we have a context space $X = \{x_{dc}, x_{bt}, x_{wv}, x_p, x_t, x_u, x_w, x_v\}$ and we must define for each of this dimension a contribution function.

The definition of these contribution functions is critical because they determine how much each attribute affects the vulnerability of an area and therefore the decisions made by the operator. Operators therefore often want to have control over these functions, so that even if the index calculation is automatic once defined, they are the ones who have determined what constitutes a risk situation. In our case, the contribution functions and weights associated with each attribute were defined in an expert-based manner, i.e., decided by domain experts. There is nothing to prevent contribution functions from being defined using other approaches, such as data-driven ones. In this case, however, even if the definition of these functions would be faster, this approach takes control away from experts, which could create resistance in sensitive areas such as the maintenance of critical.

In any case, the use of approaches based on contribution functions has the enormous advantage of being explainable: at any time, it is possible to trace

the reasons that led to a choice, understanding why a risk situation relating to one area was considered more critical than another, or vice versa. This aspect is fundamental in the context of CHCI protection.

For the sake of simplicity, in this proof of concept we define such contribution functions as independent from the zones and therefore the same for all, although in our methodology we provided that they could also be made ad hoc for each zone. We also present a simplified version of the contribution functions. A more detailed description will be provided in Chapter 7, during the testing of the complete implementation of the system. Here we define them only for illustrative purposes, in view of the subsequent proof of concept.

The contribution functions $con(x_{dc})$ (Eq. 4.12), $con(x_{bt})$ (Eq. 4.13) and $con(x_{wv})$ (Eq. 4.14) are related to the number of maintenance issues detected in the last available image of a zone. As the number of identified issues increases, the contribution made by the function to the vulnerability index increases, so that the area is being targeted as more vulnerable.

$$con(x_{dc}) = \begin{cases} 0 & x_{dc} = 0, \\ 0.2 & x_{dc} \in (0, 2], \\ 0.6 & x_{dc} \in (2, 5], \\ 0.8 & x_{dc} \in (5, 10], \\ 1 & x_{dc} \in (10, +\inf]. \end{cases} \tag{4.12}$$

$$con(x_{bt}) = \begin{cases} 0 & x_{bt} = 0, \\ 0.4 & x_{bt} \in (0, 1], \\ 0.6 & x_{bt} \in (1, 3], \\ 0.8 & x_{bt} \in (3, 5], \\ 1 & x_{bt} \in (5, +\inf]. \end{cases} \tag{4.13}$$

$$con(x_{wv}) = \begin{cases} 0 & x_{wv} = 0, \\ 0.2 & x_{wv} \in (0, 5], \\ 0.4 & x_{wv} \in (5, 10], \\ 0.6 & x_{wv} \in (10, 30], \\ 1 & x_{wv} \in (30, +\inf]. \end{cases} \tag{4.14}$$

The contribution functions $con(x_p)$ (Eq. 4.15), $con(x_t)$ (Eq. 4.17), $con(x_u)$ (Eq. 4.18) and $con(x_w)$ (Eq. 4.20) are related to weather context variables. In the case of the precipitation x_p, the contribution is affected by the mm of precipitation per hour, with the contribution obviously rising the higher the rainfall strength. In our work, the contribution function $con(x_p)$ (Eq. 4.15) is obtained from the literature [164] [165], modifying the range also on the basis

of ρ (Eq. 4.16), which takes into account the presence of other maintenance damaged conduits or tiles on roofs. Even high or low temperatures x_t in °C or their rapid variations can negatively affect the preservation of a structure within an archaeological site. Considering humidity x_u, high humidity can cause mold and water infiltration, while low humidity instead can create a bad environment for pictorial ornaments and frescoes: in such a case μ (Eq. 4.19) is used to take into account also the presence of damaged conduits. Wind x_w on the other hand, if too strong, can cause roofs and conduits to break completely. In this case, for the contribution function we use the famous Beaufort scale [166]. The parameter ω (Eq. 4.21) takes into account the presence of damaged conduits and roofs.

$$con(x_p) = \begin{cases} 0 & x_p = 0, \\ 0.2 & x_p \in (0, 2.5 - 2\rho], \\ 0.5 & x_p \in (2.5 - 2\rho, 10 - 5\rho], \\ 0.8 & x_p \in (10 - 5\rho, 50 - 20\rho], \\ 1 & x_p \in (50 - 20\rho, +\inf]. \end{cases} \tag{4.15}$$

where:

$$\rho = \left(0.7\frac{x_{bt}}{3} + 0.3\frac{x_{dc}}{5}\right), \tag{4.16}$$

$$con(x_t) = \begin{cases} 0.1 & x_t \in [-\inf, -5], \\ 0.2 & x_t \in (-5, 0], \\ 0.1 & x_t \in (0, 10], \\ 0.0 & x_t \in (10, 28], \\ 0.2 & x_t \in (28, 32], \\ 0.3 & x_t \in (32, 35], \\ 0.4 & x_t \in (35, 37], \\ 0.5 & x_t \in (37, 42], \\ 1 & x_t \in (42, +\inf]. \end{cases} \tag{4.17}$$

$$con(x_u) = \begin{cases} 0.3 & x_u \in [0, 20 - 2\mu], \\ 0.2 & x_u \in (20 - 2\mu, 40], \\ 0.0 & x_u \in (40, 60], \\ 0.6 & x_u \in (60, 80 - 5\mu], \\ 0.8 & x_u \in (80 - 5\mu, 90 - 5\mu], \\ 1.0 & x_u \in (90 - 5\mu, 100 - 10\mu], \end{cases} \tag{4.18}$$

$$\mu = \frac{x_{dc}}{5}, \tag{4.19}$$

$$con(x_w) = \begin{cases} 0.0 & x_w \in [0, 5.4], \\ 0.1 & x_w \in (5.4, 7.9 - 2\omega], \\ 0.2 & x_w \in (7.9 - 2\omega, 13.8 - 4\omega], \\ 0.5 & x_w \in (13.8 - 4\omega, 17.1 - 6\omega], \\ 0.6 & x_w \in (17.1 - 6\omega, 20.7 - 8\omega], \\ 0.7 & x_w \in (20.7 - 8\omega, 24.4 - 12\omega], \\ 1 & x_w \in (24.4 - 12\omega, +\inf]. \end{cases} \tag{4.20}$$

$$\omega = \left(0.5\frac{x_{bt}}{3} + 0.5\frac{x_{dc}}{5}\right), \tag{4.21}$$

The presence of visitors x_v in the areas can also definitely influence the vulnerability index, as indicated by the contribution function $con(x_v)$ (Eq. 4.22). In fact, surely if there is any problem in a zone that daily has a high number of visitors, the vulnerability of the structure must go up, both because such visitors could further damage the structure and because the damage could be dangerous to the structure itself. In this case, the presence of broken tiles, as symbolized by the parameter λ (Eq. 4.23), can be used to take into account the influence of such type of maintenance issues in the contribution of visitors.

$$con(x_v) = \begin{cases} 0 & x_v = 0, \\ 0.1 & x_v \in (0, 12500 - 5000\lambda], \\ 0.2 & x_v \in (12500 - 5000\lambda, 25000 - 5000\lambda], \\ 0.3 & x_v \in (25000 - 5000\lambda, 37000 - 10000\lambda], \\ 0.4 & x_v \in (37000 - 10000\lambda, 50000 - 20000\lambda], \\ 0.5 & x_v \in (50000 - 20000\lambda, 62500 - 30000\lambda], \\ 0.6 & x_v \in (62500 - 30000\lambda, 75000 - 40000\lambda], \\ 0.7 & x_v \in (75000 - 40000\lambda, 87500 - 50000\lambda], \\ 0.8 & x_v \in (87500 - 50000\lambda, 100000 - 60000\lambda], \\ 0.9 & x_v \in (100000 - 60000\lambda, 112500 - 70000\lambda], \\ 1 & x_v \in (112500 - 70000\lambda, +\inf]. \end{cases} \tag{4.22}$$

$$\lambda = \frac{x_{bt}}{3}, \tag{4.23}$$

Each of these contributions is weighted by:

$$W = \{w_{dc}, w_{bt}, w_{wv}, w_p, w_t, w_u, w_w, w_v\}$$

which is the vector of all the weights for each contribution. Both the contribution functions and the weights used were established with the consultation of

structural experts in the field of CH. The following are the weights assigned for each contribution, which reflect the importance of each of such contribution in the $CHVI^d$ (Eq. 4.10) we must calculate.

- $w_{dc} = 0.15$;
- $w_{bt} = 0.2$;
- $w_{wv} = 0.05$;
- $w_p = 0.2$;
- $w_t = 0.05$;
- $w_u = 0.1$;
- $w_w = 0.15$;
- $w_v = 0.1$.

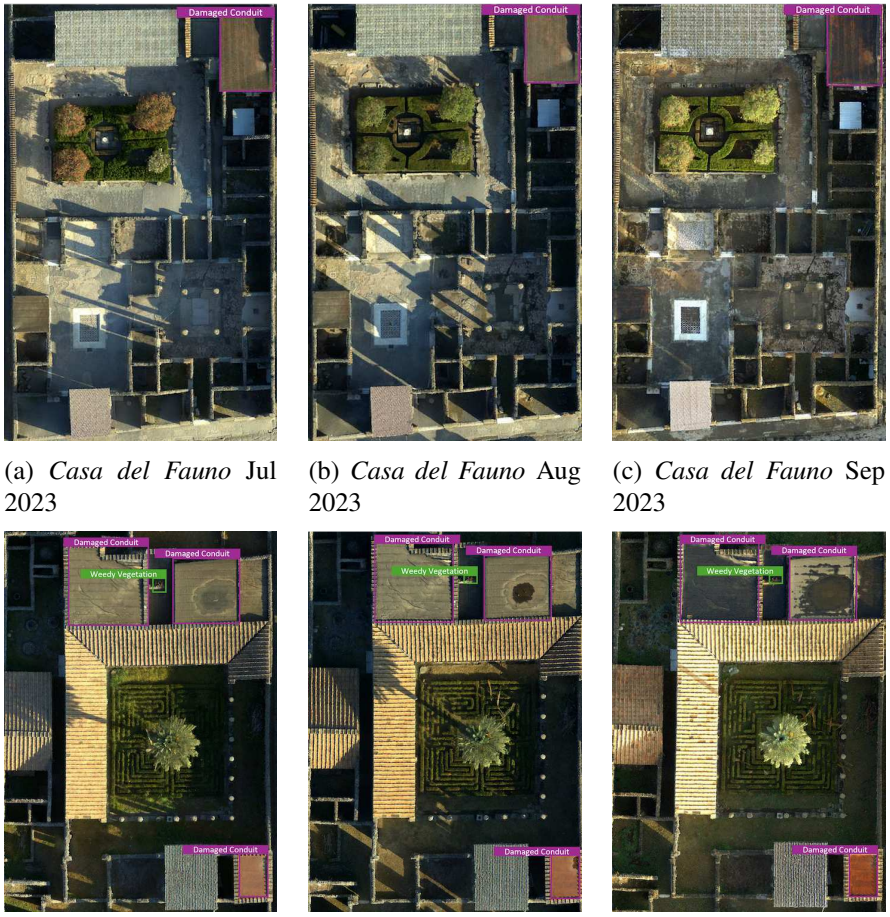
4.4.2 Case Study 1: Archaeological Site of Pompeii

The Pompeii archaeological site preserves the remains of the ancient city, buried under ash during the AD 79 eruption of Mount Vesuvius, covering an extension of 66 hectares.

The case study of the archaeological site of Pompeii is based on UAV images acquired by a high-definition RGB camera Zenmuse L1 mounted on a drone DJI 300 RTK Matrix. The drone takes photos of the entire site and then such images are used in Metashape to obtain a single georeferenced orthophoto, which generally has a size of about 10 GB, depicting the entire site. The orthophoto has a uniform ground sampling distance across the area, with the Monte Mario / Italy Zone 2 reference geo-referencing system. A manual inspection of the orthophoto when searching for maintenance issues may require a long time, with the risk of making mistakes. Too much time to identify and fix an issue could be dangerous for the health of the site: for example, if there is a damaged roof, rain could damage the frescoes. For this reason, the intervention to fix such issues needs to be sped up.

In our case study, we focused on some of the zones of the site, considering both visitable and not visitable structures. The months considered in our experimentation are July, August and September 2023: the September photos were collected after a heavy rainfall. The selected zones are *Casa del Fauno* and *Casa del Labirinto* (Figure 4.3), which show damage of varying severity to flat-roof conduits and the presence of weedy vegetation, and *Casa di Gavius*

Rufus and *Casa di Nave Europa* (Figure 4.4), which show tile damage; *Casa di Gavius Rufus* also presents weedy vegetation.



(a) *Casa del Fauno* Jul 2023

(b) *Casa del Fauno* Aug 2023

(c) *Casa del Fauno* Sep 2023

(d) *Casa del Labirinto* Jul 2023

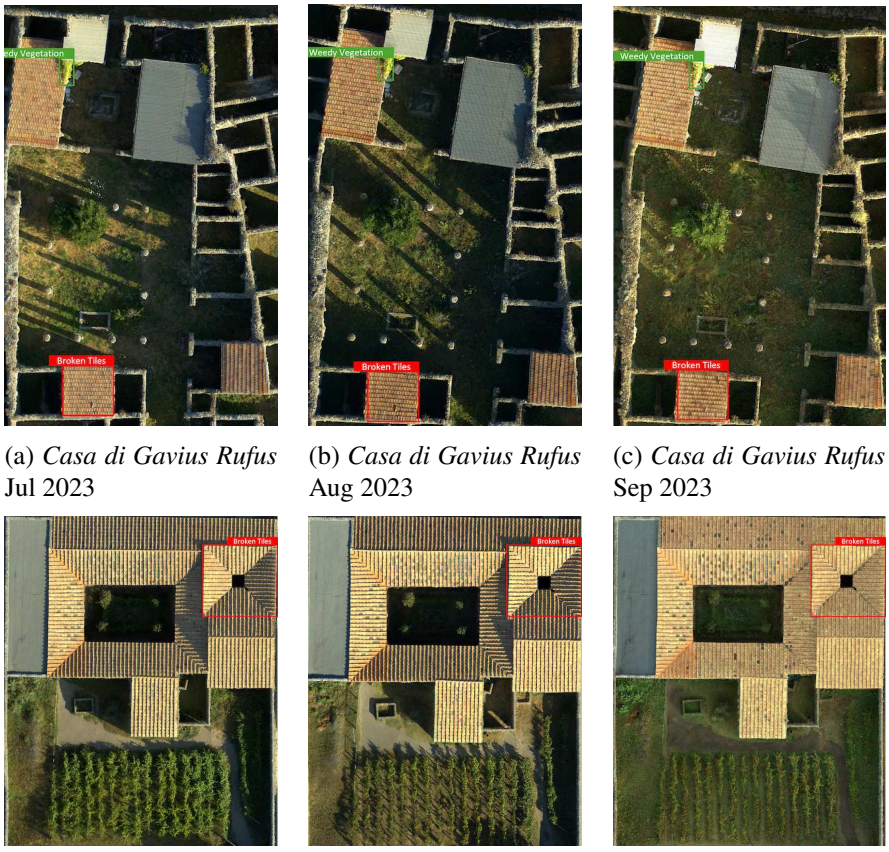
(e) *Casa del Labirinto* Aug 2023

(f) *Casa del Labirinto* Sep 2023

Figure 4.3: Images of *Casa del Fauno* and *Casa del Labirinto*.

Looking at information freely available at `https://pompeiisites.org/`, we have seen that *Casa del Fauno* and *Casa di Nave Europa* are visitable while *Casa del Labirinto* and *Casa di Gavius Rufus* are not directly visitable.

We used three orthophotos of Pompeii (one for each month), the meteorological data of the area of Pompeii for each period using Open-Meteo API, in particular data on temperature, precipitation, relative humidity, wind speed. Based on information made public by the Pompeii administration, the site recorded about 112,100 visitors per week in July, 114,100 in August, and 110,100 in September (for the visitable structures). On each UAV image, we



(a) *Casa di Gavius Rufus* Jul 2023

(b) *Casa di Gavius Rufus* Aug 2023

(c) *Casa di Gavius Rufus* Sep 2023

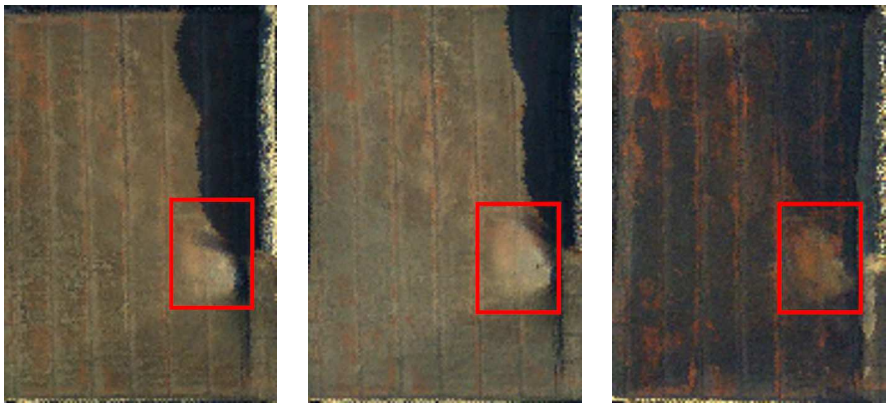
(d) *Casa di Nave Europa* Jul 2023

(e) *Casa di Nave Europa* Aug 2023

(f) *Casa di Nave Europa* Sep 2023

Figure 4.4: Images of *Casa di Gavius Rufus* and *Casa di Nave Europa*.

perform detection applying YOLOV7, which executes *L1.1 low-level perception*, *L1.2 high-level perception* and *L2.1 Situation Detection*. Trained on the images coming on various orthophotos, the model finds class and position of the issue in each of the three areas. From the images we took, our model identified a damaged conduit on *Casa del Fauno*, three damaged conduits and a weedy vegetation on *Casa del Labirinto*, a damaged tile roof on *Casa di Gavius Rufus* and another on *Casa di Nave Europa*: these damages are identified on each of the three photos because such issues have not been resolved between July, August and September.



(a) *Casa del Fauno* Jul 2023

(b) *Casa del Fauno* Aug 2023

(c) *Casa del Fauno* Sep 2023

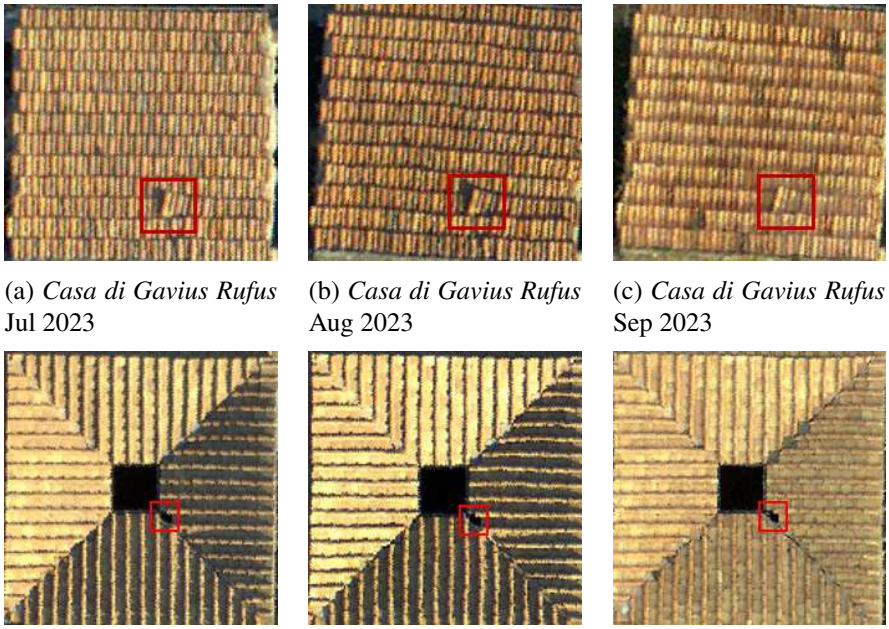
(d) *Casa del Labirinto* Jul 2023

(e) *Casa del Labirinto* Aug 2023

(f) *Casa del Labirinto* Sep 2023

Figure 4.5: Damaged Conduits on *Casa del Fauno* and *Casa del Labirinto*, (Red rectangles Highlight the Damage).

In Figure 4.5 we reported the most severe damages for each zone subject of this case study. A small damage can be found on *Casa del Fauno* in all three months (a), (b) and c, while in *Casa del Labirinto* an initial stain and seepage on the conduit (d), worsens significantly in (e) and (f). Indeed, considering that the last photo (f) was taken after a day of heavy rain, precisely on September



(a) *Casa di Gavius Rufus* Jul 2023 (b) *Casa di Gavius Rufus* Aug 2023 (c) *Casa di Gavius Rufus* Sep 2023

(d) *Casa di Nave Europa* Jul 2023 (e) *Casa di Nave Europa* Aug 2023 (f) *Casa di Nave Europa* Sep 2023

Figure 4.6: Broken Tiles of *Casa di Gavius Rufus* and *Casa di Nave Europa* (Red Rectangles Highlight the Damage).

23, and considering the already damaged conduits, this damage exposes the structure to severe mold and humidity infiltrations. In Figure 4.6, the small damage on a tile of *Casa di Gavius Rufus* remains the same in the various photos (a), (b) and (c) while in *Casa di Nave Europa* there is a missing tile. That tile remains broken in photos (d), (e), and (f), and considering the heavy rain that occurred before photo (f), such a situation could be dangerous. At this point, considering that we are in the module *L3.1 Situation Prediction*, the system has detected maintenance issues from the images obtained, and such info has been fused with contextual data in the module.

The first day of July, during the initialization of the system, since data from the perception phases are not yet available, we have only the $CHVI^s$, as in the barplot in Figure 4.7:

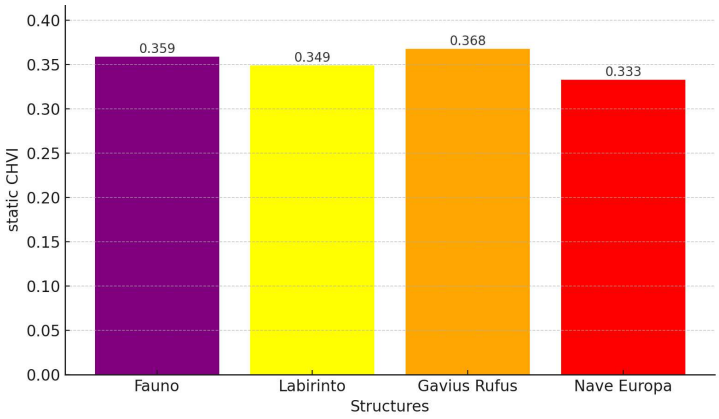


Figure 4.7: Barplot of the Standalone Static CHVI Contribution in July 2023

then the system begins to work by sensing data, going on to update the vulnerability indices over time by including the dynamic part as well. In July, the system acquires images related to the site using the UAV described in the introduction to Section 4.4, allowing us to calculate the CHVIs. The CHVIs are calculated once a day, using contextual information that is predictable up to one week in advance, so that the results are reasonably reliable, and referring to the data detected in the most recent available image. In particular, for weather data, the vulnerability indices are calculated using the maximum values of precipitation (mm per hour) and wind speed, and the maximum and minimum values of temperature and humidity. Based on these, one can decide whether to compute the CHVI using the maximum or the minimum, depending on which represents the most severe condition. For visitors, the number of people present in specific areas is estimated from the visitor counts recorded over the course of that week. The CHVIs calculated after the arrival of the orthophoto on the first day of August, and then on September 1, are reported in the barplot in

Figure 4.8.

These calculations are made by looking at the weather and visitor forecast 7 days following the day of calculation, so in this case the first few days of the month. These CHVIs, whose values over the months are always increasing, allow us to anticipate warnings and worsening of the conditions of the structures that are only noticeable in the next photo, as in the *Casa del Fauno* and *Casa del Labirinto*, where the infiltration of humidity is always getting worse over the months (Figure 4.5). From the CHVIs in both July and August, we can see how the contribution of visitors weighs a lot, as the highest contributions are on *Casa del Fauno* and *Casa di Nave Europa*, despite the fact that in the other two zones there are more vulnerabilities. During the first week of both August and September, there were no particularly severe atmospheric phenomena, except for a wind speed of up to 13.8 m/s. The last picture from September was taken after a heavy rainfall of September 23, with precipitation of 14.7 mm per hour and a maximum wind speed of 17.0 m/s. In such a context, it is fundamental to solve the issues on *Casa del Labirinto* and *Casa di Nave Europa*. In fact, even from images 4.5 and 4.6 regarding those areas, we can see the photo from August already showed heavy flooding in the stained conduits of *Casa del Labirinto*, while a hole in the tiles like the one in *Casa di Nave Europa* could damage frescoes and artifacts.

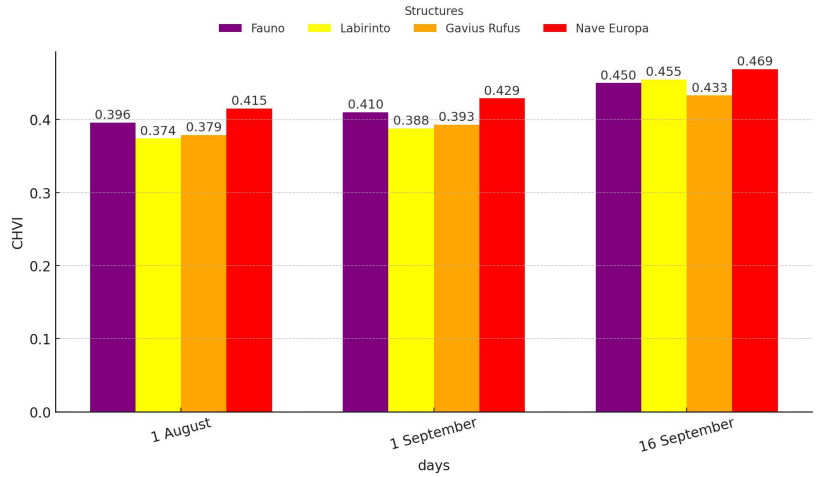


Figure 4.8: Barplot of the Pompeii CHVI Structures Across the Months of 2023

Looking at September 16, the situation regarding *Casa del Fauno* and *Casa del Labirinto* has changed, as the CHVI of the latter overtakes the other. In fact, the highest CHVIs are *Casa del Labirinto* and *Casa di Nave Europa*. With this system, we would actually be able to attend to this event, realizing the danger

as early as September 16 (a week before the event) prioritizing the areas that were actually affected by major issues. Normally, the next photo would have been available around the first day of October, leading to delays in recognizing the problem.

4.4.3 Case Study 2: Archaeological Site of Paestum

(a) *Tempio di Hera* Mar (b) *Casa Sacerdoti* Mar (c) *Altare di Nettuno* Mar

(d) *Tempio di Hera* May (e) *Casa Sacerdoti* May (f) *Altare di Nettuno* May

Figure 4.9: Satellite Images of Paestum 2024 (Arcgis, Google Earth Pro).

Paestum is an ancient city of Magna Graecia, and its ruins can be seen visiting the related archaeological site. The site of Paestum is a hybrid of structures of Greek and Roman era. The temples are one of the most ancient Greek structures arrived at the present day, while most of the structures are Greek but underwent remodelling in the Roman Republican and Imperial ages. The materials and styles used are therefore highly varied and complex, which makes maintenance activities more challenging. Compared to Pompeii, this site is more rural and features the remains of structures that are poorly preserved, open to the elements, and lacking roof coverage. This case study is conducted on three structures, in particular the *Tempio di Hera*, *Casa dei Sacerdoti*, *Altare di Nettuno*. We retrieved two aerial satellite photos of the structures, taken in March and May 2024 respectively. The images are shown in Figure 4.9, where the images (a) and (d) are related to the *Tempio di Hera* respectively in March and May, then (b) and (e) are the *Casa dei Sacerdoti* and finally (c) and (f) are

related to the *Altare di Nettuno*. The images (a), (b) and (c) are from Arcgis while the images (d), (e), (f) are from Google Earth Pro. From these images, it can be seen that, between the two months, the maintenance issues found are different:

in March no weedy vegetation can be detected and the contribution to the vulnerability indices is purely static. However, when the May aerial image arrives, the CHVIs are updated:

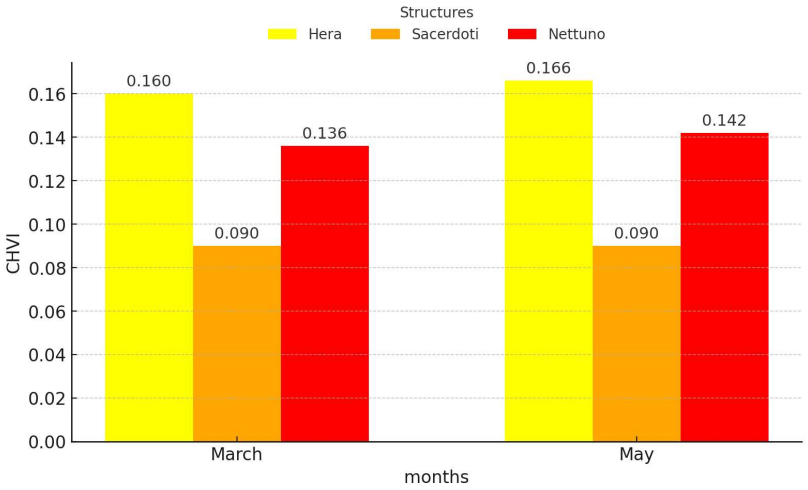


Figure 4.10: Barplot of the Paestum CHVI Structures in 2024

Since with such types of satellite images we are not able to identify ruptures or anything else, and since Paestum does not even make monthly data available on visits to structures, we only consider the contribution of weedy vegetation in the $CHVI_d$.

As can be seen in the Figure 4.11, in the *Tempio di Hera* two examples of weedy vegetation have been identified while in the *Altare di Nettuno* there are three examples. For the *Casa dei Sacerdoti* there are no identifiable weedy vegetation.

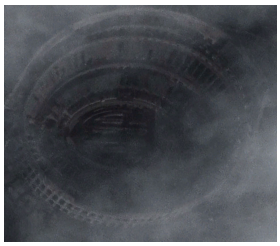
In the barplot in Figure 4.10, there is a slight increase in the CHVIs for the Tempio di Hera and the Altare di Nettuno; however, the overall perception of the site's vulnerability it is not affected since the contribution in terms of weedy vegetation it is not so high. Nevertheless, as we will also observe in the Rome case study, when the presence of weedy vegetation is more pronounced, the system is still capable of detecting meaningful changes in the vulnerability of the structures.

(a) Details of the *Tempio di Hera*'s Weedy Vegetation.

(b) Details of the *Altare Tempio di Nettuno*'s Weedy Vegetation

Figure 4.11: Paestum Weedy Vegetation May 2024.

4.4.4 Case Study 3: Archaeological Site of the Colosseum



(a) *Colosseo* Jan

(b) *Basilica* Jan

(c) *Museo Foro* Jan

(d) *Colosseo* Apr

(e) *Basilica* Apr

(f) *Museo Foro* Jan

Figure 4.12: Satellite Images of *Foro Romano* 2023 (Google Earth).

For the case study of the area around the archaeological site of the Colosseum, in Rome, we focus on three structures, in particular the *Colosseo*, *Basilica di San Clemente*, *Museo del Foro Romano*. Such structures have different characteristics from the previous case studies. The *Colosseo* is a huge ancient

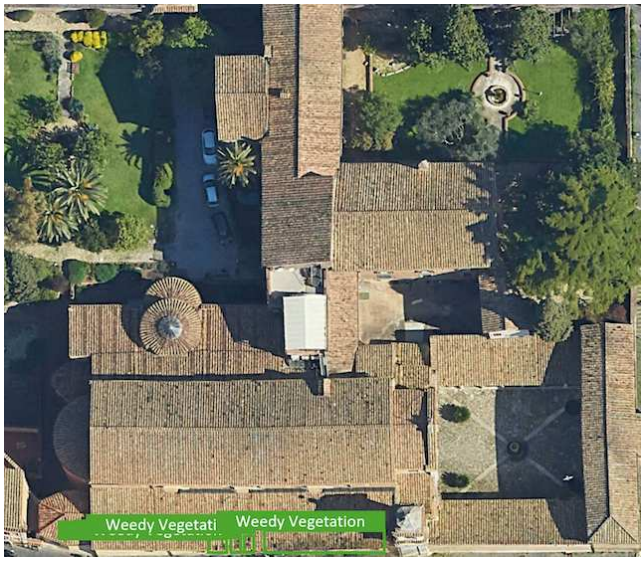
structure embedded and integrated in a modern urban environment. In such a case study, we also considered much more modern structures such as the *Museo del Foro Romano*, dating from the 20th century and the *Basilica di San Clemente*, which is a 12th-century structure that stands on the ruins of an early Christian basilica. With Google Earth Pro, we acquired photos from January and April 2023, as can be seen in Figure 4.12. In particular, the images (a) and (d) have as their subject the *Colosseo*, (b) and (e) the *Basilica di San Clemente*, (c) and (f) show the *Museo del Foro Romano*. The images (a), (b) and (c) were captured in January 2023 while (d), (e) and (f) were captured in April 2023. In January, the aerial images are completely covered by clouds, and in fact do not allow the identification of issues. Under these circumstances, the CHVIs calculated for January, visible in the barplot in Figure 4.14, have only the static contribution.

Next, we have aerial photos from April 2023, in which with regard to the Basilica and the Museum, as can be seen from the figure 4.13, it was possible to detect weedy vegetation from the tiled roofs of the two structures. In particular, in the image (a) four weedy vegetation issues were detected in the *Basilica di San Clemente* and eleven in the *Museo del Foro Romano*. The new CHVIs are also visible in Figure 4.14

The contribution related to weedy vegetation reverses the focus that operators must have on the structures at the site. Indeed, from the satellite photos, no weedy vegetation is detected on the *Colosseo*, that appears to be well maintained, at least for the low quality satellite photo we can use. Weedy vegetation has been detected for the other two structures with tile roofs. In particular, there is many samples of weedy vegetation in the *Museo del Foro Romano*'s roof, which increases the CHVI of that structure, surpassing the *Colosseo*'s CHVI. However, it is evident that in an environment of this kind, high-resolution imagery, such as that provided by UAVs in the Pompeii case study, could have enabled an improved vulnerability analysis.

4.4.5 Discussion

The three case studies show the usefulness of our approach in supporting decision-making of human operators in CH contexts. The approach leads to different results depending on the CH entities under examination and the data we use. To have good results, the method must have high quality aerial images available, so that any type of damage can be detected. This is possible with images from UAVs, since higher quality satellite images are generally not freely available and the frequency of their acquisition is lower. This explains why the system works best for our Pompeii case study, but is still useful in other



(a) Details of the *Basilica San Clemente*'s Weedy Vegetation.

(b) Details of the *Museo Foro Romano*'s Weedy Vegetation

Figure 4.13: *Foro Romano, Rome* Weedy Vegetation Apr 2023.

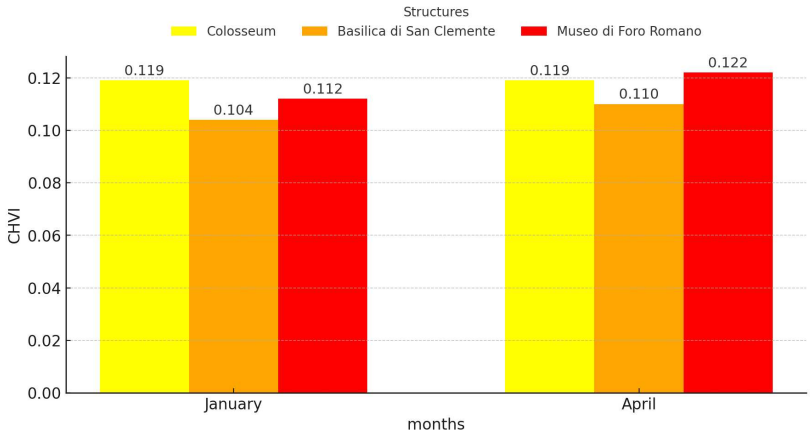


Figure 4.14: Barplot of the Colosseum CHVI Structures in 2023

contexts.

4.5 Summary

This chapter introduced the concept of a cultural heritage asset as a Cyber-Physical-Social System and highlighted the importance of supporting the situation awareness of the operators responsible for its maintenance. We modeled situations within CH assets by defining them in terms of the vulnerability of the structures to be protected. We then designed and proposed a Situation-aware Cyber-Physical-Social System methodological approach for cultural heritage protection, describing the layers and modules that compose the architecture. The system computes vulnerability indices by combining detected situations of interest with contextual information, including anthropogenic flows and weather forecasts.

We presented proofs of concept for the archaeological sites of Pompeii, Paestum, and the Colosseum. By computing both static and dynamic CHVIs from CST outputs and contextual data on weather conditions and visitor flows, the proposed framework supports human operators in identifying the structures that require prioritised interventions.

Since this chapter has presented the overall approach, the following chapters detail the proposed methods and the implementation of the modules defined in the architecture shown in Figure 4.1.

Chapter 5

Sensors and Sensing

"It is a capital mistake to theorize before one has data."

— Sherlock Holmes (Arthur Conan Doyle), *Scandal in Bohemia (1891)*

Within a CPSS, the data acquisition phase is particularly important. The control of a physical object or process requires continuous monitoring, which is essential for understanding what is happening to the object and how to act on it in order to control it and resolve any problems that may arise. For CHCI, such procedures are equally crucial: the decisions we make based on acquired data guide our actions, and if the need for maintenance on a structure goes unnoticed, this can lead to damaged structures and artifacts. The CPSS we designed includes an acquisition and pre-processing phase, with the modules highlighted in Figure 5.1.

The acquisition module relies on both physical and virtual devices from which we collect raw data about the physical object and its surrounding environment. This data can be heterogeneous: for example, images captured by UAVs or ground drones, video streams from surveillance cameras, as well as measurements from weather sensors, either virtual (via APIs) or physical (from weather stations installed directly on-site).

The collected data is often not immediately suitable for use and requires preprocessing operations, such as cleaning or reconstructing low-quality or missing data. These operations can generally be performed online, automatically, or through more complex transformations that require human intervention and must be done later, offline. In our case, the offline phase consists of generating large orthophotos from raw images captured by UAVs at the Pompeii archaeological site, whereas the online phase focuses on data imputation techniques for reconstructing missing values. The data imputation technique we propose, called GARDA, will be described later in section 5.3.2.

Once the two sensing sub-phases, offline and online, are completed, the data is passed on to the perception and then the comprehension modules, which will be described in Chapter 6.

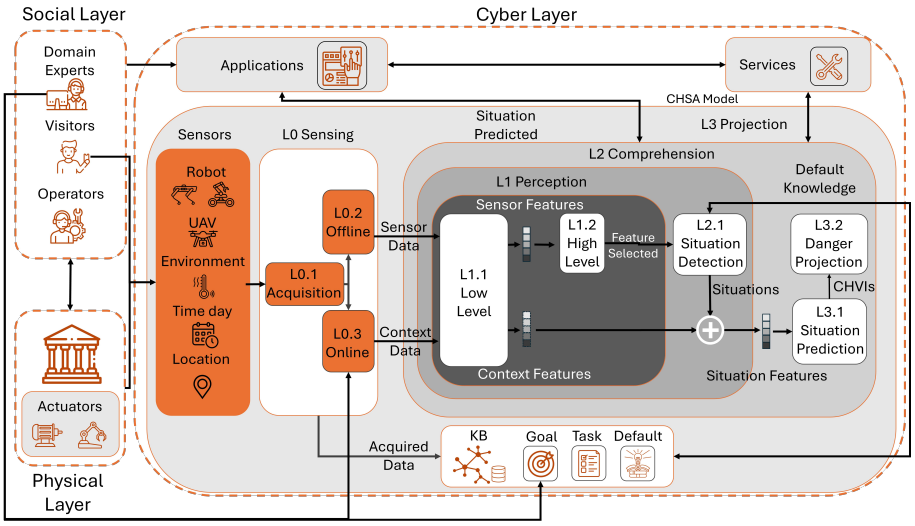


Figure 5.1: SA-CPSS with Sensors and Sensing Details

Parts of this chapter have been published in:

- Francesco Colace, Massimo De Santo, Rosario Gaeta, Rocco Loffredo. *Detection of maintenance issues from UAV images of archaeological sites: A yolo-based tool.* Multimedia Tools and Application. 2025. Volume 84. `https://doi.org/10.1007/s11042-025-20773-7`.

5.1 Acquisition Sensing

In the acquisition phase it is essential to collect data from the CHCI but also from the environment, such as weather and visitors. Let's consider the archaeological site of Pompeii, in which we have evaluated the CPSS (see Chapter 7). In such a case, the data needed are the following:

- **Aerial Images of the Site:** Aerial photographs can be captured at regular intervals, or on demand, to provide an overview of the site and to identify potential maintenance issues;

- **Contextual Information:** Real-time data on weather conditions and visitor flows within the site.

5.1.1 Aerial Images

The acquisition of aerial photographs is carried out through monthly flyovers using a DJI 300 RTK Matrix drone equipped with a high-definition RGB Zenmuse L1 camera. The DJI Matrice 300 RTK (`https://optron.com/dji/wp-content/uploads/2020/05/ds_dji-matrice-300-rtk.pdf`) combined with the Zenmuse L1 (`https://optron.com/dji/wp-content/uploads/2020/10/ds_zenmuse-l1.pdf`) forms an integrated aerial surveying system featuring a high-precision IMU, and a 1-inch CMOS RGB camera on a 3-axis stabilized gimbal. The IMU and RTK module provide centimeter-level georeferencing. The RGB camera provides accurate color capture and supports photogrammetry. Table 5.1 specify the sensors characteristics.

IMU / RGB Camera / Platform	**Specification**
IMU attitude precision	Roll/Pitch $\pm 0.025°$, Yaw $\pm 0.08°$
Camera sensor	1-inch 20 MP CMOS, mechanical shutter
Gimbal	3-axis stabilized
Positioning	RTK-enabled GNSS, centimeter-level
Protection rating	IP54 (payload)

Table 5.1: IMU, RGB Camera and Platform Specifications

The flights are carried out at an altitude of 80 meters, which, combined with the camera and the drone's capabilities, ensures high-quality surveys. To generate a complete map of the site, experts define the acquisition parameters, deciding which areas to cover, the required level of detail, the number of photographs, and the shooting angles. Based on these inputs, the drone automatically calculates the optimal flight path to minimize time and energy consumption, then follows it to capture all the necessary images. The collected photographs are subsequently processed and analysed.

5.1.2 Contextual Data

Contextual weather data is retrieved using the Open-Meteo API (`https://open-meteo.com/`). In particular, we collect information on temperature,

humidity, rainfall, and wind. Access to both historical records and forecasts is crucial to determine when specific situations require immediate attention.

Open-Meteo provides historical weather information using reanalysis datasets rather than relying on raw measurements from individual weather stations. Reanalysis is a method in which past observations coming from many different sources such as ground stations, satellites, radar, buoys and aircraft are combined with advanced numerical weather models to reconstruct the state of the atmosphere at regular intervals. This approach ensures spatial and temporal consistency. Typically the model captures large-scale patterns reliably while small-scale local variations may be smoothed out. Because the underlying models assimilate real observations, temperature and humidity tend to be highly reliable and usually differ only slightly from readings a nearby weather station would provide [167]. Wind values are also reasonably accurate although very localized events can be underestimated [168]. Precipitation remains the most challenging variable because widespread or long-lasting rainfall is usually well represented while short and intense showers that occur in very small areas may be missed or averaged out by the model [169]. In areas like Pompeii, where observational coverage is good, the reanalysis-based reconstructions perform well compared to more remote or poorly monitored regions.

From the rainfall values, it is also possible to compute the cumulative rainfall R^c, which is a fundamental piece of information for understanding the actual vulnerability of a zone, since it effectively captures the impact of rainfall over time. The value of R^c is calculated as follows:

$$R_t^c = e^{-\frac{\Delta t}{\tau}} R_{t-\Delta t}^c + R_t \tag{5.1}$$

with $\tau = 8$, and R_t rainfall value at time t. The equation must be understood as a discretized formulation with a time step $\Delta t = 1$, since the data collected from the APIs are provided on an hourly basis.

Furthermore, using data publicly provided by the Pompeii administration (`https://pompeiisites.org/parco-archeologico-di-pompei/dati-visitatori/`), we can estimate daily and weekly visitor flows to the archaeological site. Knowing which areas are accessible allows us to assess their vulnerability, since it is clear that sections in poor condition should be closed both to ensure visitor safety and to prevent damage to the site's overall image.

5.2 Offline Sensing

Once the images have been collected by the UAV, they are processed to generate a comprehensive georeferenced orthophoto of the entire site. This step requires

the active involvement of experts and dedicated photogrammetry software. In our case, Agisoft Metashape was used, a professional tool for 3D photogrammetric image processing (`https://www.agisoftmetashape.com/`). Metashape can reconstruct 3D models, point clouds, and orthophotos from digital images by applying advanced photogrammetry algorithms, a technique that enables accurate measurements and three-dimensional reconstructions of objects, terrain, and buildings.

In practice, all images are first imported into Metashape to generate a sparse point cloud through photo alignment. Control points are then added to improve georeferencing accuracy, leading to an orthophoto georeferenced into EPSG:3004 reference system (Monte Mario / Italy Zone 2). A dense point cloud is subsequently created, and the orthomosaic is produced and exported as a GeoTIFF file, as shown in Figure 5.2. At the end of the process, each orthophoto is typically around 10 GB in size.

Figure 5.2: Orthophoto of the Pompeii Site

Subsequent pre-processing steps are carried out using QGIS (`https://qgis.org/`), a free and open-source Geographic Information System (GIS) designed for creating, editing, visualizing, analyzing, and sharing geospatial data. Within QGIS, the orthophoto is imported and processed to generate layers that delineate the areas of interest, in our case the roofs. These layers are then exported as `.gpkg` files, which are used to crop the relevant sections of the orthophoto for further processing with AI algorithms.

Specifically, while the experiment described in Chapter 6 was conducted on a dataset consisting of images collected over several months and covering the entire archaeological site, the final system described in Chapter 7 was, for simplicity, applied only to Regio I. This portion of the area was extracted from the full orthophoto using polygonal vector cropping, resulting in the orthophoto

shown in Figure 5.3. The cropped orthophoto was then integrated into the final tool presented in Chapter 7.

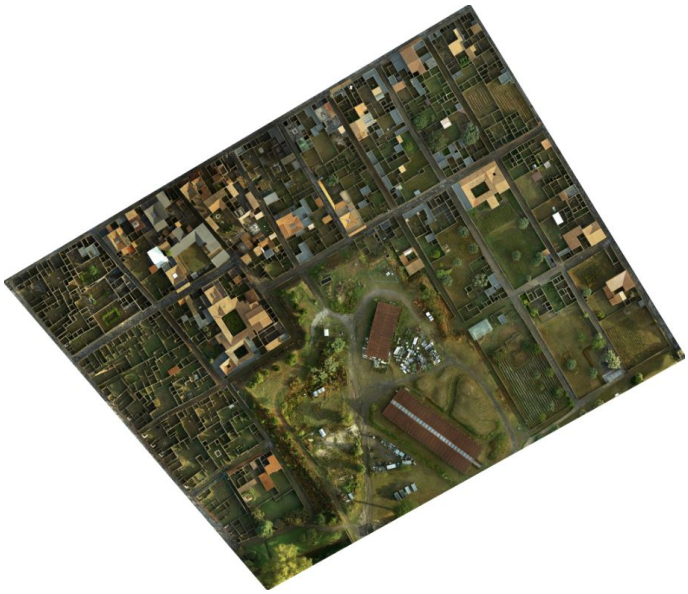


Figure 5.3: Orthophoto of the Regio I

5.3 Online Sensing

Since aerial imagery and contextual data, like the ones related to weather, may be affected by low quality or missing records, it is often necessary to reconstruct such values through a process named *Data Imputation*.

Data imputation is critical in context- and situation-aware systems [170] [171], as well as in applications that rely on data collected by sensor networks [172] [173], since missing or low-quality data are quite common [174]. Sensors and their communication networks, especially in IoT scenarios, are inherently characterized by low processing power, limited storage capacity, restricted battery life, and constrained communication bandwidth [161]. These characteristics may degrade the quality of the data collected during the perception phase, with a negative impact on the applications connected to such sensor networks [175] [176].

Data imputation is essential when analysing such data with ML techniques, since high-quality input is required to achieve reliable inference during both training and deployment and to avoid the garbage-in–garbage-out phenomenon [177]. In real-world scenarios, however, it is uncommon to have perfect datasets

without errors or missing values [178]. One possible approach to address this issue is to remove the corrupted records, although more sophisticated data imputation techniques have been shown to improve the performance of ML tasks.

Traditional imputation algorithms, such as mean, median, and mode imputation, linear regression, and k-means–based methods, can be used to recover data in such situations [179]. However, when the rate of missing data is high, these techniques may not perform well. DL approaches, on the other hand, can offer better performance, but in IoT and edge computing scenarios they are often impractical because devices frequently lack the storage capacity and computational power needed to run such models [180].

To address these challenges, this work proposes GARDA, the Granular Association-rule-based Data Imputation Approach, a novel lightweight method based on association rules and the principles of granular computing, designed for IoT devices with limited power and battery resources and for small, highly correlated sensor networks. The goal is to improve the performance of traditional approaches while requiring significantly less storage and computational capacity than DL methods. Granular computing (GrC) is a computational paradigm useful for handling uncertain data [181] [182] [183]. In this work, we exploit its principles to intelligently discretize datasets with continuous values, thereby enabling the use of association-rule-based data mining, which is typically applicable only to categorical datasets. GARDA exploits the granulation process to create intervals around the available values in the datasets. Subsequently, granular association rules are mined from such intervals in order to reconstruct the missing data by exploiting recurrent patterns in the datasets.

This approach has been integrated into the CPSS system described in Chapter 4, enabling the automatic reconstruction of missing data, particularly weather-related data, and supporting the reconstruction of drone imagery.

5.3.1 Related Works

Data imputation is the process of reconstructing missing data in a dataset [184] [185]. The distribution of missingness inside a dataset is generally classified into: i) missing completely at random (MCAR), when the missingness is unrelated to its or any other missing or observed values [186]; ii) missing at random (MAR), when the missingness is related to the observed data but not to the missing ones [187]; iii) missing not at random (MNAR), when the missingness is related to the missing data themselves [188].

We briefly review some traditional association-rule and neural network-based imputation techniques, together with state-of-the-art granular

association-rule imputation techniques.

5.3.1.1 Traditional imputation techniques

The earliest approaches to data imputation rely on replacing missing values with the mean, mode, or median of the non-missing entries in each column of a dataset. These methods perform reasonably well when data are MCAR, when more than 10% of the data are missing, or when the correlation among variables is high [189] [190]. Other traditional imputation techniques are based on regression, such as Linear Regression (LR) [191], Polynomial Regression [192], Singular Value Decomposition (SVD) [193], K-Nearest Neighbour (KNN) [194], K-Means [195], and Fuzzy C-Means [196].

5.3.1.2 Association-rule Imputation

Association-rule mining is a family of techniques used to identify recurrent patterns among items in a dataset. State-of-the-art algorithms in this field include Apriori [197], Apriori TID [198], Apriori Hybrid [199], and FP-Growth [200].

Wu et al. proposed a nominal data imputation method based on ranking association rules [201]. In a subsequent work [202], the same authors introduced a weighted voting strategy, also based on association rules, to reconstruct missing data in a dataset. Jiang [203] proposed an imputation method grounded in association-rule mining of closed frequent itemsets, that is, subsets of frequent patterns without redundancy. Chaurasia presented a technique that combines multiple imputation with rules derived from univariate F- and Beta-statistics computed over multiple imputed datasets [204]. In [205], Hong and Wu introduced an iterative missing-value completion method that uses the support values of Robust Association Rules (RARs) to extract useful rules for inferring missing entries. These techniques, however, were designed exclusively for categorical values. D'Aniello et al. extended this approach to make it applicable to IoT datasets containing continuous values [175]. A quality-aware sensor data management framework is proposed the users select the required quality levels, and virtual sensors are used to satisfy those requirements. Our work builds on the algorithm introduced in [205] and later extended to continuous values in [175], integrating these approaches with the granular computing paradigm. In particular, the proposed integration eliminates the need for manual discretization of continuous values by representing them as interval sets.

5.3.1.3 Neural network-based Imputation

Neural networks (NNs) can be used to reconstruct missing data in datasets. Many of them are designed for multivariate time-series data. In this context, there are two main categories: Recurrent-Neural-Network-based (RNN) models and Self-Attention-based models. M-RNN [206] is a multidirectional recurrent deep neural network composed of an interpolation block, which exploits correlations within each data stream, and an imputation block, which exploits correlations across data streams. SAITS [207] is based on self-attention and uses a weighted combination of two DMSA (diagonally masked self-attention) blocks to learn both temporal and feature correlations among the measurements.

5.3.1.4 Granular imputation

Data imputation is inherently uncertain because, once a piece of information is lost, it is impossible to recover it with complete certainty. The intrinsic ability of granular computing to handle uncertain and noisy data makes it well suited for data imputation tasks.

Granular Computing (GRC) has emerged as a conceptual and algorithmic platform aimed at the representation and processing of information granules [208] [209]. Granules are the fundamental building blocks of granular computing, where a granule is any subset, class, object, cluster, or element of a universe that can be seen as a more specific and restricted part of it [210]. Granulation aims at the composition and decomposition of granules [211]: they can be fuzzy sets, shadow sets, rough sets, and interval sets. In this work, we will use the interval set [212], in which an element is part or not of a granule if its value is contained within the boundaries of the interval representing the granule. Along with the set theory comes a well-developed discipline of interval analysis [213]. According to [209], an interval set A is described in the following way:

$$A(x) = \begin{Bmatrix} 1, & if\ x \in A \\ 0, & if\ x \notin A \end{Bmatrix} \tag{5.2}$$

where $A(x)$ stands for a value of the characteristic function of set A at point x.

In particular, Zhong et al. proposed a two-stage imputation mechanism, in which the first phase performs numerical imputation and then constructs a granular representation of the missing values [214]. In their work, imputation is first applied to individual variables in the dataset and is then refined through the use of information granules. A second approach adopts fuzzy clustering, specifically Fuzzy C-Means (FCM), in order to identify structural patterns within the data and subsequently use this information to improve the imputation process.

Hu et al. proposed an innovative classification method for incomplete data that relies on information granules [183]. The first part of their study aims to reveal the structural backbone of several labeled data subspaces by applying fuzzy clustering to the missing values, while the second part focuses on building information granules that reconstruct and represent the missing values using the refined prototypes and the insights obtained from the classification process.

Hu et al. proposed a method that forms information granules to represent missing data and then constructs granular fuzzy models directly from the hybrid data obtained by combining these granules with the available numerical values [182].

5.3.2 GARDA: Granular Association-rule-based Data Imputation Approach

The proposed Granular Association-Rule-Based Data Imputation Approach (GARDA) is shown in Figure 5.4. GARDA consists of three phases: (i) granulation, which discretizes continuous values in the incomplete dataset; (ii) the use of granular association rules to partially impute missing values by identifying the intervals in which they may lie; and (iii) a traditional data imputation technique (for example, linear regression) to fill in the remaining missing values. Each of these three phases is described in the pseudocode of Algorithm 1.

The motivation behind this hybrid approach lies in the fact that the granular rule-based method, which builds on the discrete imputation strategy proposed by Hong et al. [205], performs very well when the rules have high confidence. However, when rule confidence is low, its imputation performance deteriorates and becomes worse than that of traditional data imputation techniques. Traditional methods, on the other hand, are outperformed by the rule-based approach on the cells where the latter imputes with high confidence. For this reason, we decided to propose an approach that combines the strengths of both techniques. Furthermore, once some of the missing cells have been imputed using the most reliable rules of the granular method, the subsequent traditional imputation step can operate under more favorable conditions and achieve a more accurate reconstruction. These considerations are supported by experimental evidence, which will be presented in the following subsections.

The GARDA approach, illustrated in Figure 5.5, starts from a dataset represented as a matrix in which some cells are empty. The missing values in these cells are reconstructed by exploiting recurrent patterns among the dataset attributes and leveraging the correlations between their values.

The dataset is processed as shown in Figure 5.4 by the following three phases:

1. Granulation: the procedure used to discretize continuous values. The values are grouped into granules whose characteristics depend on the specific granulation method adopted. For columns that already contain categorical data, this step can be skipped;

2. GrC Association-rule-based Imputation: It takes as input an incomplete granular dataset in which each value is represented by a granule and may contain missing information. The approach uses only the most reliable and confident rules extracted from the underlying patterns. These rules are applied to fill a subset of the originally missing values in the dataset. The continuous values are then reconstructed from such granules, which is a step not required for columns that originally contained categorical data;

3. Traditional Imputation: The eventually remaining missing data are reconstructed by a traditional imputation algorithm (e.g., linear regression).

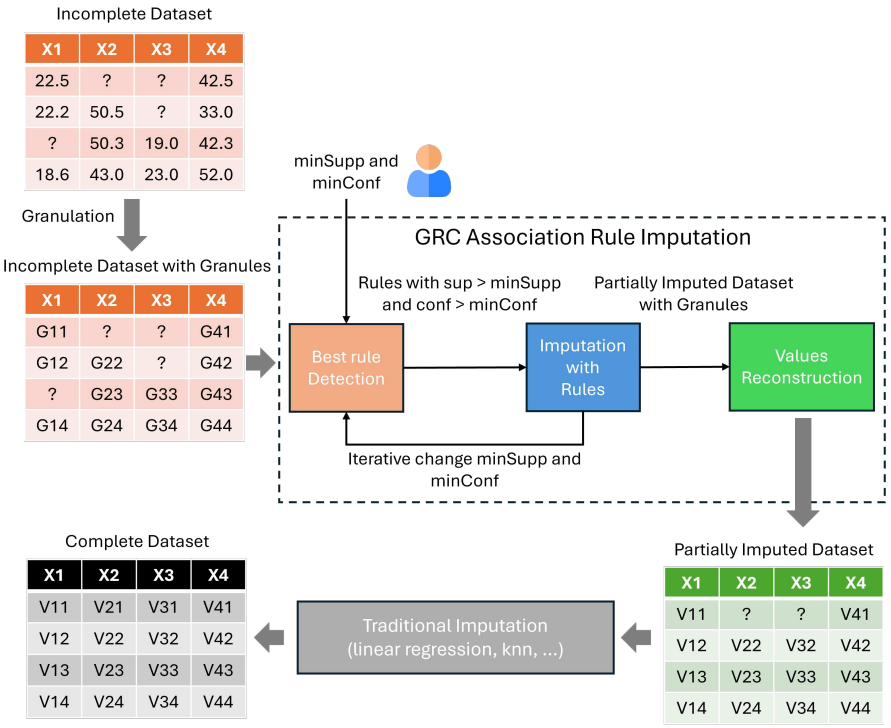


Figure 5.4: GARDA Architecture

5.3.2.1 *Formal definition of the approach*

Let $\chi = \{x_1, x_2, x_3, ..., x_n\}$ be a set of variables with $x_i \in \mathbb{R}$ and $1 \leq i \leq n$ and $\mathcal{D}$ be a dataset which is a multiset of a subset of χ. We define a family of granulation function $\Gamma_i(x_i, k_i)$ in which the specific function applied depends on x_i and k_i is the maximum number of granules relative to x_i. In particular, the function $\Gamma_i(x_i, k_i) : x_i \in \chi \rightarrow \{g_{ij}\}$, with $1 \leq j \leq k_i$, will output k_i granules for each x_i. In this way, each value of x_i is associated to granule g_{ij}.

Let $\mathcal{I} = \{g_{11}, g_{12}, ..., g_{21}, g_{22}, ..., g_{ij}\}$ be the set of granules generated, called items. A p-itemset X_p is a subset of the set of granules I containing p items, i.e., $X_p \subseteq \mathcal{I}$ with $p = |X_p|$. Considering S a multiset of subsets (a dataset of already formed granules) of $\mathcal{I}$, we define $\sigma(X_p)$ as the transaction set for an itemset X_p as follows:

$$\sigma(X_p) = \{T | X_p \subseteq T, T \in S\} \tag{5.3}$$

where T represents a single transaction in the dataset S.

Considering two generic itemsets X_p and $Y_{p'}$ with p and p' the cardinalities of the itemsets (where typically $p' = 1$), an association rule can be defined as an expression $X_p \rightarrow Y_{p'}$ with $X_p \cap Y_{p'} = \varnothing$.

If the set S contains missing values, it is convenient to exploit the concept of Robust Association Rule (RAR) proposed by Ragel and Cremilleux [215] [216]. The RAR approach is designed to mine association rules in an incomplete dataset. Let $Dis(X_p)$ be the set of the disabled (missing) data of the itemset X_p:

$$Dis(X_p) = \{T | \exists A \in X_p, A = ?, X_p \subseteq T, T \in S\} \tag{5.4}$$

where ? is the missing value, and A is the missing attribute belonging to X_p. The RAR support for an itemset X_p can be defined as:

$$Sup(X_p) = \frac{|\sigma(X_p)|}{|S| - |Dis(X_p)|} \tag{5.5}$$

The confidence for an association rule $X_p \rightarrow Y_{p'}$ based on the RAR approach is defined as follows:

$$Conf(X_p \rightarrow Y_{p'}) = \frac{|\sigma(X_p \cup Y_{p'})|}{|\sigma(X_p)| - |Dis(Y_{p'}) \cap \sigma(X_p)|} \tag{5.6}$$

Based on such definitions, we design GARDA, with its flow diagram depicted in Figure 5.5. Each phase of the approach is described in the next sections.

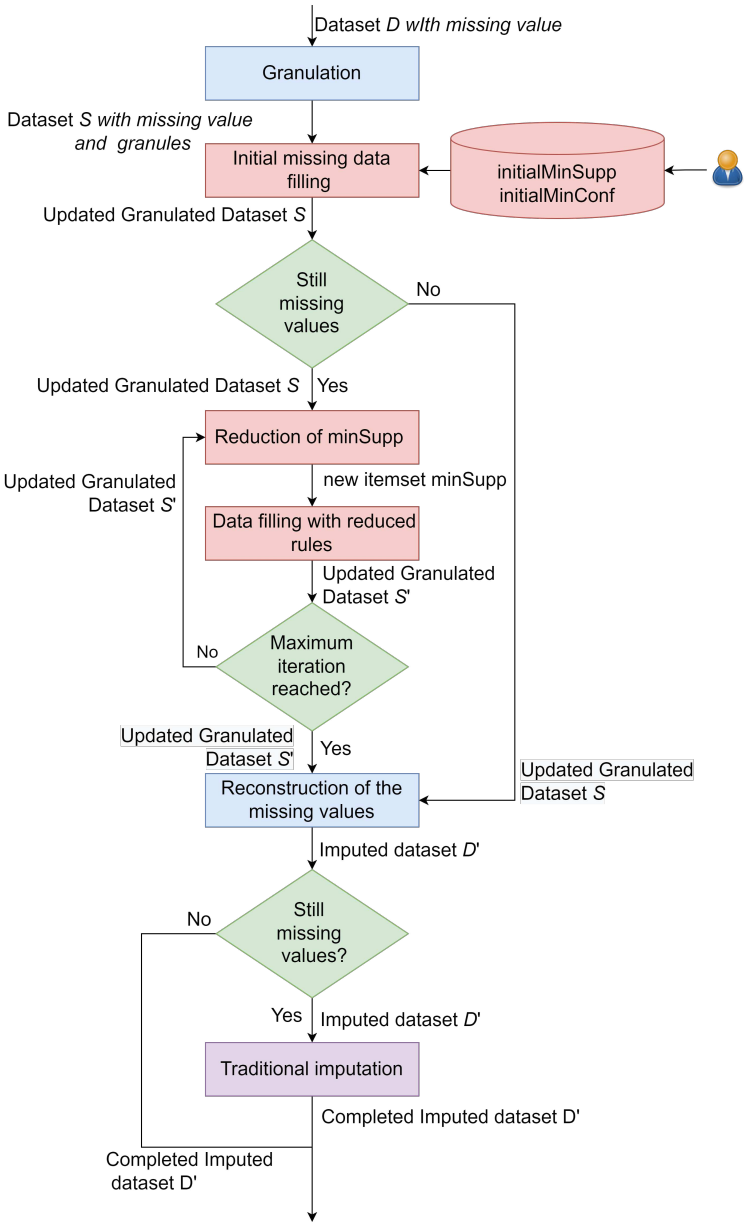


Figure 5.5: GARDA Flow Diagram

5.3.2.2 Granulation approach

GARDA begins with an incomplete dataset and aims to fill all missing values, whether discrete or continuous. An automatic granulation method is employed to generate the information granules, avoiding the need for manual discretization as proposed, for example, in [175]. Naturally, if the values are already discrete, the granulation step can be skipped.

In this section, we define three types of unsupervised granulation approaches: Equal-Width [217], a static method; K-Means [218], a clustering-based approach; and Kernel Density Estimation [219], a probabilistic method. Figure 5.6 illustrates the results of granulation using the Equal-Width method [217]. These approaches define the granulation functions to be applied to each type of attribute. A granulation function must be specified for each variable to be discretized.

- **Equal-Width:** It starts from the minimal value a_{i0} and the maximal value a_{ik} of the variable x_i in D, and divides the range into k_i equal-width discrete intervals. The approach calculates the interval width λ_i:

$$\lambda_i = \frac{(a_{ik} - a_{i0})}{k_i} \tag{5.7}$$

  and then granulates the continuous values into k_i discrete intervals bounded: $\big\{[a_{i0}, a_{i1}), ..., [a_{i(j-1)}, a_{ij}), ...[a_{i(k-1)}, a_{ik}]\big\}$,

$$a_{ij} = a_{i0} + j\lambda_i \tag{5.8}$$

  where $j = 1, 2, \ldots, k$.

  Eventually, the outliers are included in the granules at the edges of the range.

- **K-Means:** Another granulation approach implemented in GARDA is K-Means clustering [218]. In this approach, discretization emerges from grouping similar numerical values rather than uniformly partitioning the domain. The method applies the K-Means algorithm to the attribute values of x_i, specifying the desired number of clusters k_i. Once the centroids $c_1, \ldots, c_{k_i}$ have been computed, each value is assigned to the nearest cluster. For each resulting cluster C_j, the corresponding discretized interval is defined as:

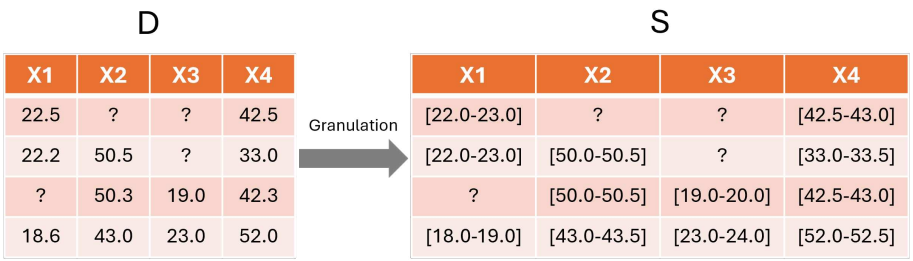
$$I_j = [\min(x_i \in C_j),\ \max(x_i \in C_j)].$$

  Such granulation approach is capable of capturing the underlying structure of the data, resulting in intervals narrower in dense regions and wider in sparsely populated ones.

- **Kernel Density Estimation (KDE):** introduced in [219], is a data-driven granulation method that adapts interval boundaries according to the estimated probability density. It identifies regions where the density changes significantly and places cut-points accordingly. For each candidate cut-point T, the density induced by binning is compared with the density estimated through a kernel density estimator, and a score is computed as:

$$\text{Score}(T) = \sum_i \left(p(x_i) - f(x_i)\right),$$

where $f(x_i)$ is the binned density and $p(x_i)$ is the kernel density estimate. Cut-points with the highest scores correspond to intervals that poorly represent the true density and are therefore refined. The process iteratively splits such intervals, and the optimal number of intervals is determined via a cross-validated log-likelihood criterion, making the method data-driven.



X1	X2	X3	X4
22.5	?	?	42.5
22.2	50.5	?	33.0
?	50.3	19.0	42.3
18.6	43.0	23.0	52.0

X1	X2	X3	X4
[22.0-23.0]	?	?	[42.5-43.0]
[22.0-23.0]	[50.0-50.5]	?	[33.0-33.5]
?	[50.0-50.5]	[19.0-20.0]	[42.5-43.0]
[18.0-19.0]	[43.0-43.5]	[23.0-24.0]	[52.0-52.5]

Figure 5.6: Application of the Granulation Approach to an Incomplete Dataset.

Figure 5.6 provides an example of the granulation mechanism, in this case using Equal-Width. In the starting D dataset, x_1 and x_3 are elements of the same type of attribute (for example, temperature), while x_2 and x_4 are elements of another type of attribute (for example, humidity). In this example, for the attributes x_1 and x_3, an interval width $\lambda_{x_1} = \lambda_{x_3} = 1$, while for x_2 and x_4, the interval width is $\lambda_{x_2} = \lambda_{x_4} = 0.5$. The granulation process produces the granular dataset S.

Once the granulation approach has replaced the non-missing values with their corresponding granules, missing data are imputed using granular association rules. Missing values for which no rules with sufficient confidence are found are left as missing. At this point, the algorithm returns a dataset that is still incomplete but with values grouped in granules, and for this reason discretized.

5.3.2.3 GrC Association-rule-based Imputation

The imputation of the missing granular intervals is performed using the granular association rules, applying the procedure reported in the Algorithm 1, which describes its whole pseudo-code. The approach used is Apriori-like: indeed, to the best of our knowledge, no data-mining approaches with lower complexity than Apriori-like methods currently exist for the imputation of missing data, precisely because algorithms that improve upon Apriori typically rely on structural information that cannot be exploited when datasets contain missing values.

In our approach, granular association rules mined from the original incomplete dataset are first used to roughly reconstruct missing values in Steps 1–4. The input to the algorithm is the dataset S, where each element is an interval-valued granule I_i. Rule mining is performed based on the initial minimum support and minimum confidence thresholds selected by the user, following the steps below:

1. Find the RAR-support values of all the 1-itemsets. If the support of a 1-itemset X_1 is not smaller than the threshold, put it in the set of frequent (large) 1-itemsets, L_1;

2. Iteratively, find the other frequent itemsets with more than one item in an Apriori-like way using the RAR-support evaluation. The set of frequent (large) k-itemsets is called L_k;

3. Find the confidence value of each possible candidate granular association rule generated from the frequent itemsets. If the confidence of such a rule is not smaller than the threshold $minConf$, it is inserted in the set of the granular association rules, GAR;

4. Use the set GAR to infer the missing values of the incomplete dataset with the following substeps.

   (a) If there is only one granular association rule which can be used to derive the missing value of an attribute in a tuple, then use the rule;

   (b) If there are more than one granular association rule which can be used (i.e., multiple matches with the antecedents of such rules), then use the one with the maximum RAR confidence value; if more than one rule have the same maximum RAR-confidence values, then use the one with the maximum RAR-support value; if the maximum RAR-support values are still the same, then keep the value still missing if the rules derive different values. The updated dataset is indicated with S'.

If at this point the dataset is still not completed, the minimum support is reduced and more association rules are found, in order to fill as many values as possible using only rules with at least a minimum confidence value, chosen by the user. This second part works iteratively, until the dataset is complete, or the minimum support and confidence become too low, or a maximum number of iterations is reached.

5. Check the dataset; if there are still missing values in the dataset, continue with the Step 6; otherwise, simply execute the reconstruction of the values, which is visible in Figure 5.5;

6. Let $y = 1$, where y is used to control the reduced minimum support value and the reduced minimum confidence threshold;

7. Set the reduced minimum support threshold $RedMinSup$, and the reduced minimum confidence threshold $RedMinConf$ where $rsup$ and $rconf$ are the reduced coefficients for support and confidence;

8. For each tuple with missing values, find the set of originally non-missing attribute-value pairs in the tuple from the granular dataset and create the candidate 2-itemsets, with one item from the set of non-missing attribute-value pairs and the other from the possible values in a missing value;

9. Find the RAR-support values of all the candidate 2-itemsets from the transformed granulated dataset S; if the RAR-support of a candidate 2-itemset X_2 is not smaller than the reduced support threshold $RedMinSup$, put it in the set of reduced frequent (large) 2-itemsets, L_2';

10. Find the confidence value of each possible candidate association rule generated from the frequent 2-itemsets, L_2', generated in Step 9. If the confidence of a candidate granular association rule is not less than the reduced confidence threshold, put it in the set of reduced granular association rules, GAR';

11. Use the set of granular association rules GAR' to infer the missing values of the updated dataset S' with the following substeps:

    (a) If there is only one granular association rule which can be used to derive the missing value of an attribute in a tuple, then use the rule;

    (b) If there is more than one granular association rule which can be used to derive the missing value of an attribute in a tuple, then use the one with the maximum RAR confidence value. If more than

Algorithm 1 GARDA

Input IncompleteDataset, InitialminSupport, InitialminConfidence, y_{MAX}, reducedCoefficient
Output ImputedDataset

▷ **Granulation**
if *IncompleteDataset* have continuous values **then**
    IncompleteDiscreteDataset ← *Granulation(IncompleteDataset, DiscretizationParameters, MinMaxPerAttributes)*
else
    IncompleteDiscreteDataset ← *IncompleteDataset*
end if
▷ **GrC Association-rule-based Imputation**
rows ← *getRows(IncompleteDiscreteDataset)*
Step 1:
for all *row* in *rows* **do**
    for all *item* in *row* **do**
        if *support(item)* >= *InitialMinSupport* **then**
            L1.add(item)
        end if
    end for
end for
Step 2:
L2 ← *get2LargeItemSets(L1)*
Lk.addAll(L2)
do
    k++
    LkList.addAll(Lk)
    createKplus1Itemset(Lk, CkPlus1, LkPlus1, k)
    Lk ← *LkPlus1*
while *size(LkPlus1)* > 0
Step 3:
for all *associationRule* from itemsets in *LkList* **do**
    if *Conf(associationRule)* $>$ *minConf* **then**
        add *associationRule* to *associationRules*
    end if
end for
Step 4:
for all *row* in *rows* **do**
    for all *item* in *row* **do**
        if only an association rule is found **then**
            use such rule
        end if
        if more association rules are found **then**
            Use the rule with maximum support or confidence value
        end if
        if more association rules are found with same support and confidence **then**
            leave the value missing
        end if
    end for
end for
Step 5 and 6:
$y \leftarrow 1$
while *missingValuesInDataSet* and $y < y_{MAX}$ and $minSupport > 0$ **do**
Step 7:
$minSupport \leftarrow \frac{InitialMinSupport}{reducedCoefficient \cdot y}$
$minCoefficient \leftarrow \frac{InitialMinCoefficient}{reducedCoefficient}$
Step 8:
    for all *row* in *rows* **do**
        for all *missingValue* in *missingValues* **do**
            for all *possibleValue* in *possibleValues* **do**
                add itemset(missingItem, possibleValue) to C2b
            end for
        end for
    end for
Step 9:
    for all *itemset* in *C2b* **do**
        if $itemsetSupport < minSupport$ **then**
            add itemset to L2b
        end if
        add itemset(missingItem, possibleValue) to C2b
    end for
Step 10:
    $reducedAssociationRules \leftarrow addRulesFromItemSets(L2b, true)$
Step 11 and 12:
    for all *row* in *rows* **do**
        for all *item* in *row* **do**
            if only an association rule is found **then**
                use such rule
            end if
            if more association rules are found **then**
                Use the rule with maximum support or confidence value
            end if
            if more association rules are found with same support and confidence **then**
                leave the value missing
            end if
        end for
    end for
    y++
end while
▷ **Reconstruction and Traditional Imputation**
if *IncompleteDataset* had continuous values **then**
    PartiallyFilledDataset ← *ValuesReconstruction(IncompletedDiscreteDataset)*
else
    PartiallyFilledDataset ← *IncompletedDiscreteDataset*
end if
if *PartiallyFilledDataset* still have missing data **then**
    CompleteDataset ← *TraditionalImputation(PartiallyFilledDataset)*
else
    CompleteDataset ← *PartiallyFilledDataset*
end if

one rule has the same maximum RAR-confidence value, then use the one with the maximum RAR-support value. If the maximum RAR-support values are still the same and the rules infer different values, then keep the missing value still unknown.

12. $y = y + 1$ and repeat Steps 7–11 until there are no missing values in the updated dataset or y gets to a predefined y_{max} value.

5.3.2.4 Value reconstruction and Traditional Imputation

At this point, the granules must be converted back into continuous values. In the case of K-Means, the continuous value is given by the corresponding cluster centroid, whereas for Equal-Width and Kernel Density Estimation, the reconstruction function used is the mean value between the two boundaries of the interval in the following way:

$$rec_i(g_{ij}) = \frac{startInterval_{ij} + endInterval_{ij}}{2} \tag{5.9}$$

An example of the reconstruction process is synthesized in Figure 5.7. Obviously, the records that were not missing in the original dataset can be

S Imputed

X1	X2	X3	X4
[22.0-23.0]	**[50.0-50.5]**	**[19.0-20.0]**	[42.5-43.0]
[22.0-23.0]	[50.0-50.5]	**[19.0-20.0]**	[33.0-33.5]
[22.0-23.0]	[50.0-50.5]	[19.0-20.0]	[42.5-43.0]
[18.0-19.0]	[43.0-43.5]	[23.0-24.0]	[52.0-52.5]

Reconstruction

D Imputed

X1	X2	X3	X4
22.5	**50.25**	**19.5**	42.5
22.2	50.5	**19.5**	33.0
22.5	50.3	19.0	42.3
18.6	43.0	23.0	52.0

Figure 5.7: Dataset Reconstructed with the Mean of the Intervals. The Initial Missing Values Reconstructed by the Algorithm are in Bold.

restored to their original value.

If, at the end of this phase, the dataset still contains missing values, a traditional imputation method is applied to complete the remaining empty records. This method can impute all values for which the confidence and quality of the rule-based imputation would have been too low, including cases where an entire row in the original dataset is missing. In such situations, the rule-based approach cannot operate because no rules can be applied. Traditional techniques, on the other hand, can handle these cases and complete the imputation, while also benefiting from the reduced number of missing values that remain to be filled.

5.3.3 Evaluation

At this point, we will test the GARDA approach to highlight any advantages it may offer. We will then comment on the datasets used, the evaluation procedure carried out, and the results obtained, including the strengths and weaknesses of our approach.

5.3.3.1 Datasets

The experiments for our approach were conducted on the Intel Lab Dataset, containing data collected from 54 sensors deployed in the Intel Berkeley Research Lab (`https://db.csail.mit.edu/labdata/labdata.html`) between February 28 and April 5, 2004. Each entry originates from a Mica2Dot sensor, which measures temperature, humidity, light, and battery voltage. Each measurement has the following format: date, time, epoch, mote ID, temperature, humidity, light, and voltage. The sensors recorded one measurement approximately every 31 seconds, producing a dataset of about 2.3 million readings from the 54 sensors installed in the lab. The epoch value is a monotonically increasing sequence number assigned to each measurement.

Considering that most of the days after March 14 contain many outlier values (mainly due to low battery voltage), we focus our analysis on the period from March 1 to March 14, creating a dataset that includes only the measurements collected during this interval. We concentrate on temperature readings. To address the fact that some sensors do not provide measurements at every time instance, we group the data into time slots of $t = 15$ minutes and compute the average temperature within each slot.

The experiments of the algorithms have been performed on a dataset containing the sensors 1, 2, 3, 4, 6, 7, 9, 10, 33, and 35. Their measurements are characterized by a high spatial correlation that can be exploited in the imputation process, as shown in the correlation matrix of Figure 5.8.

The approaches SAITS and M-RNN are compared with GARDA using the Beijing Multi-Site Air-Quality Dataset [220] (`https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data`). This dataset contains hourly air pollutant and meteorological measurements from 12 monitoring sites in Beijing, collected between 2013/03/01 and 2017/02/28 (a total of 48 months). Each row includes a row ID, date and time, the concentration in μg/m^3 of PM2.5, PM10, SO_2, NO_2, CO, O_3, and the following meteorological variables: TEMP (temperature, °C), PRES (pressure, hPa), DEWP (dew point temperature, °C), RAIN (precipitation, mm), WSPM (wind speed, m/s), and WD (wind direction). Since our method focuses on continuous variables, the dataset was reduced to TEMP, PRES, DEWP, and WSPM; the

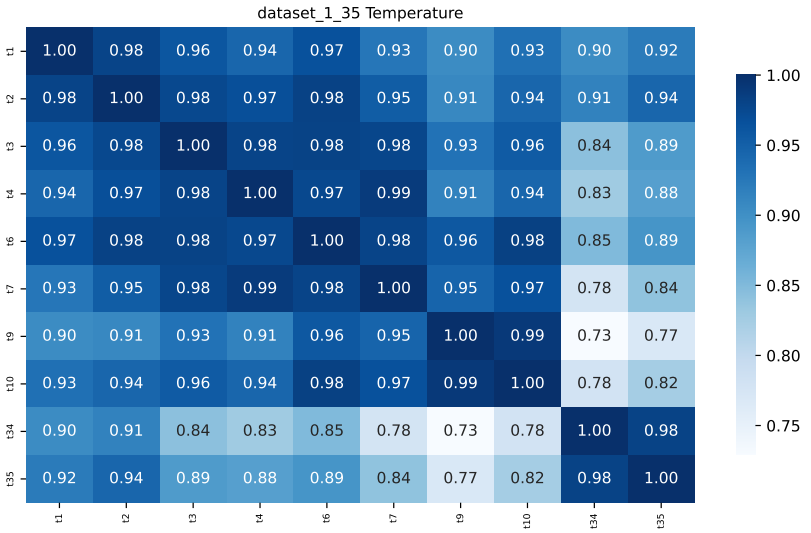


Figure 5.8: Correlation Matrix of the Ten Sensors Considered in the Intel Lab

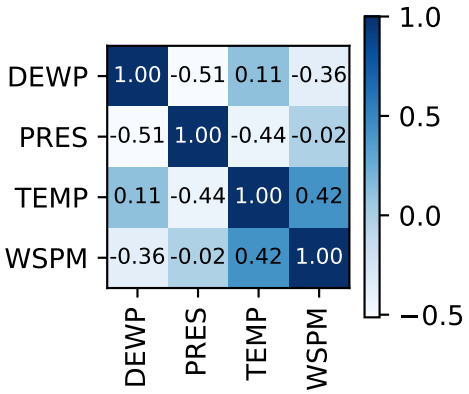


Figure 5.9: Correlation Matrix of the 4 Sensors from the Moderate Correlation Version of the AirQuality Beijing Dataset

RAIN variable was excluded because it takes the value zero for a large portion of the dataset.

Furthermore, because our method is designed to impute short time frames locally, as typically generated by IoT devices, we restricted the time frame used for imputation to 40 days and considered data from five stations (Dingling, Dongsi, Nongzhanguan, Aotizhongxin, and Guanyuan). To ensure a fair comparison, the SAITS and M-RNN approaches use this 40-day window as the test set, while the remaining data is split into validation and training sets with a 1:3 ratio. The measurements exhibit strong correlations among sensors of the same type across different stations.

To test our method on datasets characterized by low correlation, we created a reduced dataset containing four sensors from a single station (Dingling), since the highest correlations in the original dataset occur among measurements of the same type across different stations. The correlation matrix of this modified dataset is shown in Figure 5.9. For this version of the dataset, 200 days of measurements were included in order to maintain approximately the same number of data points.

5.3.3.2 Evaluation Protocol

To evaluate the ability of an algorithm to properly reconstruct the data, a ground truth is required. For this reason, the values already missing in the original dataset were not considered in our assessment, and the corresponding rows were removed. Starting from a complete dataset, we artificially inject missingness into some of the records; we then reconstruct these values and compare the imputed results with the ground truth in order to measure reconstruction accuracy. The experiments follow the evaluation protocol shown in Figure 5.10, adopting an approach commonly used in related works [175, 221, 222]. Since real-world datasets may exhibit different levels of noise and missingness, we evaluate the method under various missing rates. In particular, we generate datasets with 5%, 10%, 20%, 30%, 40%, and 50% missing values [175, 221–223]. Different error-injection models have been used. In our experiments, we simulate both MCAR and consecutive-missing patterns for the Intel Lab Dataset. For the Beijing Multi-Site Air-Quality dataset, we evaluate the method under the MCAR model only, since this is the model adopted in the corresponding paper [207], which evaluates both its methodology and the approach presented in [206].

MCAR introduces completely random errors, whereas the consecutive missingness model simulates a more realistic scenario in which a sensor, or a group of sensors, fails to produce any readings for a certain period of time (e.g., due to network errors or power outages). In our setup, this period ranges from 4 to 64 hours. Since, in the Intel Lab dataset, each row represents the average value of

the sensors over a 15-minute interval, the number of consecutive missing values for a sensor ranges from 16 to 256. To obtain results that are more reliable and less influenced by randomness or fluctuation, we generated ten different datasets for each error model and each missing rate, using different random seeds. The evaluation metrics are then computed as the average over these ten datasets for each combination of error model and missing rate. Specifically, the following metrics are considered:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{5.10}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{5.11}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5.12}$$

where y_i is the i^{th} true value, and $\hat{y}_i$ the i^{th} predicted value. Since the Air Quality dataset includes values that are close to zero, the MAPE cannot be used to measure the relative error between real and imputed values. This is because the true values appear in the denominator, resulting in excessively high MAPE values when they are near zero. The MRE metric solves this problem by calculating the ratio between the sums of the errors and the true values.

$$MRE = \frac{\sum_{i=1}^{N}|y_i - \hat{y}i|}{\sum_{i=1}^{N}|y_i|} \tag{5.13}$$

Therefore, when comparing our method to ML approaches, we use MRE instead of MAPE.

This procedure is performed to compute the metrics for all the algorithms used in the comparison, in particular: Mean, Median, Mode, Linear Regression, KNN, K-Means, and SVD-based for the Intel Lab Dataset, while we compare against M-RNN and SAITS for the Beijing Multi-Site Air-Quality dataset.

5.3.3.3 Results

The granular algorithm (Algorithm I) of the GARDA approach is configured in the following way, after a grid search which took into account each of this parameters:

- $initSup = 0.65$

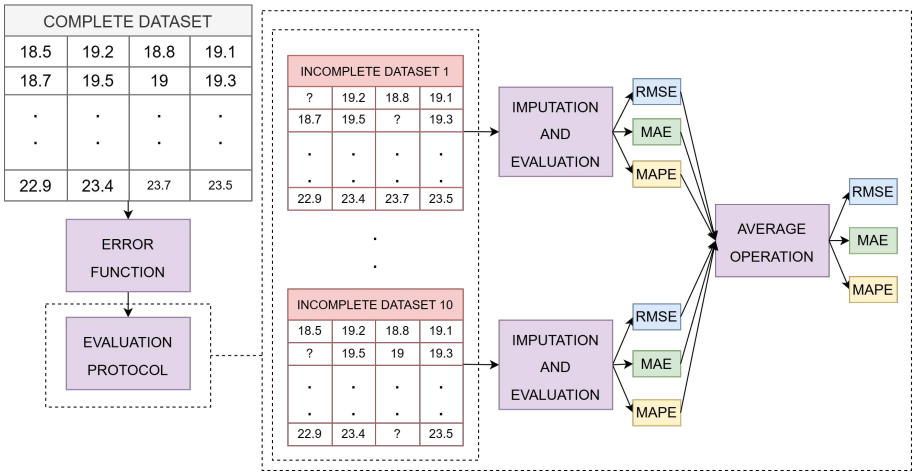


Figure 5.10: Evaluation Approach of GARDA on Various Datasets

- $initConf = 0.9$
- $rSup = 100$,
- $rConf = 1$,
- $numIter = 1$.

Other critical aspects are the chosen granulation technique and the granule size. For our implementation of GARDA, we tested Equal-Width, K-Means, and Kernel Density Estimation.

Granule size is crucial, as it affects both the number of items used in the reconstruction and the accuracy with which the continuous value is recovered from the granule. Indeed, larger granules decrease the total number of items (thus increasing support and confidence), but they also lead to greater reconstruction error, as the granule's average (or centroid) value is computed over a broader interval. On the other hand, smaller granules produce more items, with lower support and confidence, yet they enable a more precise reconstruction of the original value when the appropriate granule is chosen. For this reason, the granule size for both datasets was selected through a grid search guided by the standard deviation σ estimated from the dataset. In particular, we tested the following values: $\sigma, \frac{\sigma}{2}, \frac{\sigma}{3}, \frac{\sigma}{4}, \frac{\sigma}{5}, \frac{\sigma}{6}, \frac{\sigma}{7}, \frac{\sigma}{8}$. For the Intel Lab dataset, considering that σ calculated over the entire dataset is 2.31, the best value was obtained with K-Means, with an indicative granule size of $\frac{\sigma}{6}$ (0.385), corresponding to 41 clusters. Clearly, for K Means this value is only indicative within the grid search, because the algorithm determines the actual cluster sizes based on the data distribution. K-Means also achieved the best performance on the

Beijing dataset; the grid search was performed by starting from a fraction of σ considered reasonable for the scale of the values in the column, then progressively reducing it until reaching the first value at which performance deteriorates. The fraction of σ immediately preceding this value is then taken as the optimal choice. Following such procedure, for TEMP ($\sigma = 4.75$), we tested $\frac{\sigma}{4}, \frac{\sigma}{5}, \frac{\sigma}{6}, \frac{\sigma}{7}, \frac{\sigma}{8}$, and the best value was $\frac{\sigma}{7}$ (0.679), corresponding to 42 clusters. For PRES ($\sigma = 7.52$), we tested $\frac{\sigma}{12}, \frac{\sigma}{13}, \frac{\sigma}{14}, \frac{\sigma}{15}, \frac{\sigma}{16}$, and the best value was $\frac{\sigma}{15}$ (0.501), corresponding to 88 clusters. For DEWP ($\sigma = 6.35$), we tested $\frac{\sigma}{6}, \frac{\sigma}{7}, \frac{\sigma}{8}, \frac{\sigma}{9}, \frac{\sigma}{10}$, and the best value was $\frac{\sigma}{9}$ (0.706), corresponding to 46 clusters. For WSPM ($\sigma = 1.55$), we tested $\sigma, \frac{\sigma}{2}, \frac{\sigma}{3}, \frac{\sigma}{4}$, and the best value was $\frac{\sigma}{3}$ (0.517), corresponding to 18 clusters.

For the traditional data imputation algorithm (second phase of the GARDA architecture in Figure 1), a Bayesian ridge regression model with the following parameters is used;

- $\alpha_1 = 1e^{-6}$: shape parameter for the gamma distribution prior over the alpha parameter,
- $\alpha_2 = 1e^{-6}$: inverse scale parameter (rate parameter) for the gamma distribution prior over the alpha parameter,
- $\lambda_1 = 1e^{-6}$: shape parameter for the gamma distribution prior over the lambda parameter,
- $\lambda_2 = 1e^{-6}$: inverse scale parameter (rate parameter) for the gamma distribution prior over the lambda parameter,
- $\alpha_{init} = \frac{1}{Var(Y)}$: initial value for alpha,
- $\lambda_{init} = 1$: initial value for lambda.

These are the parameters of the other imputation algorithms compared with GARDA: mean, median, and mode approaches have no parameters to tune; the linear regression has the same parameters used in our method; KNN has a number (K) of neighbours of 10; K-Means has a K value of 20; the SVD-based approach has a shrinkage value equal to $max(\text{singular value init matrix})/50$. Such values have been tuned with a grid-search approach to find the best parameters for the considered datasets. The hyper-parameters for the ML methods are the same used by the SAITS paper [207], as they have been already optimized for the Air Quality dataset. They are set to the following:

- SAITS and M-RNN: sequence length 24, batch size 128, ADAM optimizer with 0 weight decay, early stop patience 30.

- SAITS only: learning rate 0.000882, hidden dim 512, hidden size of feed forward layer 512, self-attention number of heads 4, key dim 128, value dim 64, dropout rate 0, inner group parameter sharing with 1 group.
- M-RNN only: learning rate 0.000924, hidden size 256.

Tables 5.2 and 5.3 report the means μ and standard deviations σ obtained on ten datasets for all the metrics. Mean, median, and mode imputation have been excluded because their performances are far worse than that of the other methods. The KNN shows the lowest RMSE for missing rates below 40%, while its performance decreases for higher missing rates. In contrast, GARDA achieves the lowest RMSE for high missing rates. In particular, GARDA improves upon LR for high missing rates, with an improvement of about 20% for both the 40% and 50% missing rate. Considering the MAE, KNN performs best for low missing rates, whereas our approach achieves the lowest MAE for the 40% and 50% missing rates. In terms of MAE, GARDA improves upon LR by 15% when the error rate is 50%. More precisely, our approach yields an average error of about 0.29 to 0.33 C° on the 40% and 50% datasets, compared to 0.33 to 0.39 C° for LR.

The performance of GARDA has also been evaluated under a consecutive error model using the Intel Lab Dataset (Table 5.3). Under this model, KNN performs worse than in MCAR case.

The best performing model is GARDA, as it matches LR, which is the traditional method with the highest 5% missing rate performance, while it outperforms all other approaches from 10% on-wards. This advantage becomes increasingly evident at higher missing rate. Specifically, GARDA surpasses LR with RMSE reductions of 46% at 10% missingness (0.378 vs. 0.700), 52% at 20% (0.529 vs. 1.109), 62% at 30% (0.642 vs. 1.715), 70% at 40% (0.752 vs. 2.529), and 67% at 50% (1.206 vs. 3.761). Also considering the MAE, GARDA matches LR at 5% and outperforms all other methods at 10%, 20%, 30%, 40%, and 50% missing rates, leading, for instance, to a 71% reduction in MAE at 50% missingness with respect to LR.

Tables 5.2 and 5.3 report the performance of the Hong et al. [205] method, in which all missing values are imputed using rules rather than leaving the remaining ones to a traditional technique. The poor results arise because imputations are performed even for cells for which only low-confidence rules are available.

A summary of the granulation-method grid search on the Intel Lab dataset under the consecutive error model is presented in Table 5.4, which compares the best configurations identified for each granulation strategy. K-Means slightly outperforms Equal-Width at lower missing rates (5–20%), yielding an average reduction of about 2% in both RMSE and MAE, except for RMSE at the 20% missing-rate level. More substantial gains appear at higher missing rates: at

Table 5.2: Test results from the Missing Completely Random Error Model

Rate	Measure	KNN			LR			SVD			KMEANS			HONG et al. [205]			GARDA		
		rmse	mae	mape	rmse	mae	mape	rmse	mae	mape	rmse	mae	mape	rmse	mae	mape	rmse	mae	mape
5%	μ	**0.199**	**0.101**	**0.004**	0.301	0.189	0.008	0.456	0.282	0.012	0.516	0.367	0.016	2.274	1.608	0.071	0.299	0.190	0.008
	σ	**0.0201**	**0.0073**	**0.0003**	0.0205	0.0096	0.0004	0.0271	0.0010	0.0004	0.0270	0.0140	0.0006	0.1958	0.1026	0.0037	0.0170	0.0079	0.0004
10%	μ	**0.220**	**0.111**	**0.005**	0.326	0.202	0.009	0.488	0.297	0.013	0.528	0.372	0.017	2.348	1.660	0.073	0.320	0.200	0.009
	σ	**0.0021**	**0.0070**	**0.0003**	0.0014	0.0083	0.0004	0.0151	0.0006	0.0003	0.0252	0.0099	0.0004	0.2062	0.1260	0.0055	0.0148	0.0074	0.0003
20%	μ	**0.273**	**0.137**	**0.006**	0.378	0.227	0.010	0.523	0.317	0.014	0.534	0.380	0.017	2.153	1.535	0.068	0.353	0.218	0.010
	σ	**0.0019**	**0.0054**	**0.0002**	0.0014	0.0007	0.0003	0.0172	0.0063	0.0003	0.0155	0.0010	0.0004	0.1329	0.0869	0.0032	0.0108	0.0059	0.0003
30%	μ	**0.364**	**0.188**	**0.008**	0.470	0.272	0.012	0.580	0.350	0.015	0.551	0.385	0.017	2.220	1.579	0.069	0.422	0.258	0.011
	σ	**0.0135**	**0.0051**	**0.0002**	0.0030	0.0073	0.0003	0.0241	0.0075	0.0003	0.0153	0.0057	0.0003	0.1783	0.1097	0.0046	0.0195	0.0085	0.0004
40%	μ	0.575	0.306	0.013	0.583	0.330	0.015	0.642	0.384	0.017	0.576	0.401	0.018	2.122	1.494	0.066	**0.478**	**0.289**	**0.013**
	σ	0.0331	0.0146	0.0006	0.0027	0.0086	0.0004	0.0169	0.0039	0.0002	0.0150	0.0086	0.0004	0.1799	0.1438	0.0060	**0.0167**	**0.0091**	**0.0004**
50%	μ	0.841	0.484	0.021	0.698	0.398	0.018	0.718	0.433	0.019	0.610	0.420	0.019	1.956	1.333	0.066	**0.568**	**0.333**	**0.015**
	σ	0.0222	0.0112	0.0005	0.0042	0.0017	0.0007	0.0105	0.0041	0.0002	0.0263	0.0098	0.0004	0.0688	0.0469	0.0021	**0.0376**	**0.0134**	**0.0006**

Table 5.3: Test results from the Consecutive Error Model

Rate	Measure	KNN			LR			SVD			KMEANS			HONG et al. [205]			GARDA		
		rmse	mae	mape	rmse	mae	mape	rmse	mae	mape	rmse	mae	mape	rmse	mae	mape	rmse	mae	mape
5%	μ	0.404	0.266	0.012	**0.339**	**0.218**	**0.010**	0.423	0.298	0.013	1.086	0.664	0.028	2.178	1.703	0.076	0.345	0.218	0.010
	σ	0.0853	0.0561	0.0025	**0.0880**	**0.0500**	**0.0022**	0.1101	0.0686	0.0029	0.4299	0.2142	0.0086	0.4670	0.3733	0.0159	0.0934	0.0564	0.0024
10%	μ	0.947	0.595	0.027	0.700	0.453	0.020	0.896	0.622	0.027	2.359	1.403	0.060	2.381	1.817	0.081	**0.378**	**0.240**	**0.011**
	σ	0.2283	0.0976	0.0041	0.0682	0.0495	0.0023	0.1249	0.0679	0.0027	0.5328	0.2316	0.0087	0.6254	0.4347	0.0182	**0.0707**	**0.0522**	**0.0024**
20%	μ	1.678	1.002	0.044	1.109	0.716	0.032	1.529	1.014	0.044	3.805	2.217	0.095	2.511	1.833	0.079	**0.529**	**0.300**	**0.013**
	σ	0.2839	0.1061	0.0040	0.0560	0.0392	0.0017	0.2119	0.1016	0.0039	0.3889	0.1846	0.0071	0.4275	0.3320	0.0145	**0.2166**	**0.0720**	**0.0031**
30%	μ	2.576	1.474	0.065	1.715	1.068	0.047	2.277	1.458	0.063	5.326	3.070	0.131	2.519	1.784	0.078	**0.642**	**0.362**	**0.016**
	σ	0.2539	0.0878	0.0033	0.1573	0.0682	0.0030	0.2031	0.0889	0.0033	0.3317	0.1525	0.0057	0.3971	0.2427	0.0106	**0.1762**	**0.0672**	**0.0026**
40%	μ	3.575	2.019	0.088	2.529	1.555	0.068	3.129	1.975	0.085	6.928	4.007	0.171	2.330	1.595	0.069	**0.752**	**0.433**	**0.019**
	σ	0.1746	0.0564	0.0019	0.3110	0.1560	0.0068	0.1659	0.0811	0.0033	0.2557	0.0966	0.0036	0.3466	0.2278	0.0098	**0.1356**	**0.0620**	**0.0025**
50%	μ	5.054	2.791	0.120	3.761	2.293	0.100	4.190	2.617	0.112	8.826	5.139	0.220	2.381	1.542	0.066	**1.206**	**0.649**	**0.028**
	σ	0.2259	0.1137	0.0050	0.3318	0.1721	0.0075	0.1857	0.0929	0.0039	0.2052	0.1063	0.0047	0.3297	0.2473	0.0104	**0.2754**	**0.0953**	**0.0035**

Table 5.4: Granulation Method Test Results from the Consecutive Error Model

Rate	Measure	EQUAL-WIDTH (41 intervals)			KMEANS (41 clusters) [218]			KERNEL DENSITY [219]		
		rmse	mae	mape	rmse	mae	mape	rmse	mae	mape
5%	μ	0.348	0.222	0.0010	**0.345**	**0.218**	**0.010**	0.915	0.571	0.026
	σ	0.0895	0.0529	0.0023	**0.0934**	**0.0564**	**0.0024**	0.2230	0.1599	0.0072
10%	μ	0.379	0.242	0.011	**0.378**	**0.240**	**0.011**	0.947	0.586	0.026
	σ	0.0692	0.0515	0.0023	**0.0707**	**0.0522**	**0.0024**	0.1457	0.1127	0.0047
20%	μ	**0.522**	0.300	0.013	0.529	**0.300**	**0.013**	1.125	0.720	0.032
	σ	**0.3052**	0.1149	0.0045	0.2166	**0.0720**	**0.0031**	0.1248	0.0852	0.0036
30%	μ	0.681	0.374	0.016	**0.642**	**0.362**	**0.016**	1.210	0.786	0.035
	σ	0.2523	0.0873	0.0034	**0.1762**	**0.0672**	**0.0026**	0.1427	0.0884	0.0036
40%	μ	0.856	0.464	0.020	**0.752**	**0.433**	**0.019**	1.440	0.948	0.041
	σ	0.1923	0.0756	0.0029	**0.1356**	**0.0620**	**0.0025**	0.2339	0.1263	0.0051
50%	μ	1.284	0.676	0.028	**1.206**	**0.649**	**0.028**	1.784	1.139	0.049
	σ	0.3221	0.1147	0.0044	**0.2754**	**0.0953**	**0.0035**	0.2448	0.1265	0.0053

Table 5.5: Test Results against Neural Network-based Methods - High Correlation Version

Rate	GARDA			M-RNN			SAITS		
	RMSE	MAE	MRE	RMSE	MAE	MRE	RMSE	MAE	MRE
5	0.191	0.076	0.094	0.404	0.259	0.374	**0.167**	**0.055**	**0.080**
10	**0.185**	0.074	**0.091**	0.450	0.285	0.404	0.185	**0.065**	0.093
20	**0.183**	**0.074**	**0.091**	0.492	0.330	0.473	0.229	0.083	0.119
30	**0.190**	**0.076**	**0.093**	0.514	0.350	0.502	0.258	0.100	0.143
40	**0.195**	**0.078**	**0.095**	0.541	0.375	0.537	0.278	0.110	0.158
50	**0.215**	**0.083**	**0.101**	0.574	0.400	0.573	0.301	0.125	0.179

Table 5.6: Test Results against Neural Network-based Methods - Moderate Correlation Version

Rate	GARDA			M-RNN			SAITS		
	RMSE	MAE	MRE	RMSE	MAE	MRE	RMSE	MAE	MRE
5	0.685	0.521	0.641	0.774	0.626	0.689	**0.333**	**0.167**	**0.188**
10	0.703	0.534	0.659	0.810	0.654	0.731	**0.323**	**0.172**	**0.193**
20	0.726	0.554	0.686	0.840	0.680	0.766	**0.333**	**0.172**	**0.194**
30	0.756	0.579	0.716	0.844	0.715	0.805	**0.345**	**0.181**	**0.203**
40	0.784	0.602	0.745	0.905	0.737	0.833	**0.350**	**0.182**	**0.205**
50	0.812	0.626	0.776	0.933	0.763	0.86	**0.368**	**0.199**	**0.224**

30% missingness, RMSE and MAE decrease by 5% and 3%, respectively; at 40%, by 12% and 7%; and at 50%, by 6% and 4%.

Table 5.5 reports the metrics calculated on the high correlation version of the Beijing Air Quality dataset. SAITS shows better performances across all metrics at 5% and 10% missing rate, but from 20 percent upwards GARDA outperforms SAITS, reaching a 29% reduction in RMSE (0.215 vs. 0.301) and 34% in MAE (0.083 vs. 0.125) at 50% missing rate.

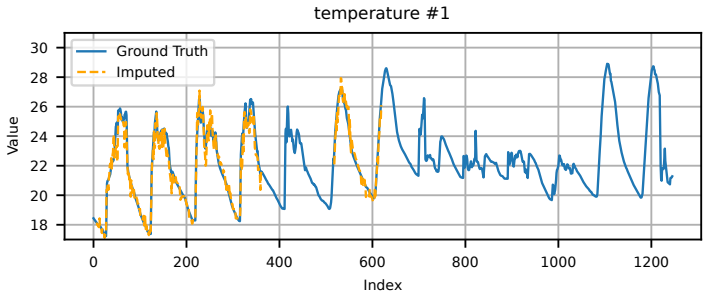
Finally, Table 5.6 reports the metrics for the low-correlation dataset, where SAITS outperforms all other methods across all missing rates. However, GARDA performs better than M-RNN. Moreover, GARDA maintains good performance while being suitable for devices with low computational power and limited storage, unlike neural network approaches.

5.3.3.4 Discussion

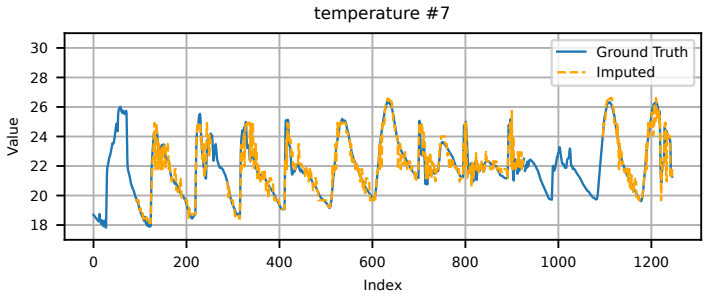
Such experimental phase yields promising imputation results, outperforming both traditional and machine learning methods in many context. The experiments reveal clear strengths and limitations, helping to identify the scenarios in which the method excels. In particular, GARDA delivers substantial improvements at missing rates above 10% when the dataset exhibits strong correlations. GARDA is specifically designed for integration into IoT-based systems that rely heavily on edge computing environments in which only small, low-power, battery devices with limited computational and storage resources are available.

The GARDA performance results reported in Tables 5.2, 5.3, 5.5, and 5.6 were obtained using K-Means as the granulation method, as a grid search indicated that it outperforms both Equal-Width and Kernel Density Estimation. The grid search on the Intel Lab dataset shows that K-Means consistently achieves better results across all metrics and missing rates, with only a minor exception: at a 20% missing rate, RMSE are slightly better with Equal-Width. This behaviour can be attributed to the greater flexibility of K-Means in determining granule size, unlike Equal-Width, which constrains all granules to a fixed width. Thanks to this flexibility, K-Means can form narrower clusters in denser regions of the data and wider clusters in sparser ones. As a result, rules with the highest confidence correspond to more frequent values and narrower intervals, leading to more accurate reconstructions.

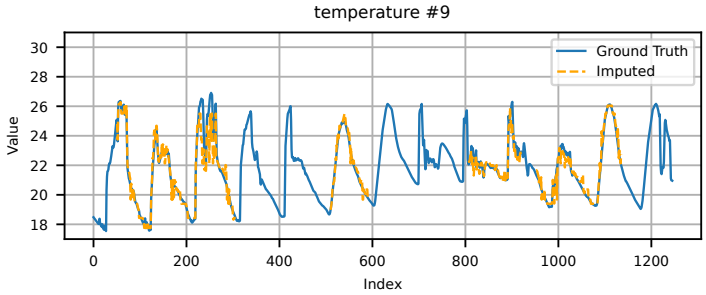
Furthermore, GARDA's improvements for MCAR models over traditional methods at high missing rates align well with our initial expectations. The core idea behind GARDA come from observing that, while traditional methods are computationally lightweight, their performance degrades significantly as the missing rate increases. Our goal was therefore to design a hybrid approach that integrates rule-based and traditional techniques to enhance overall imputation



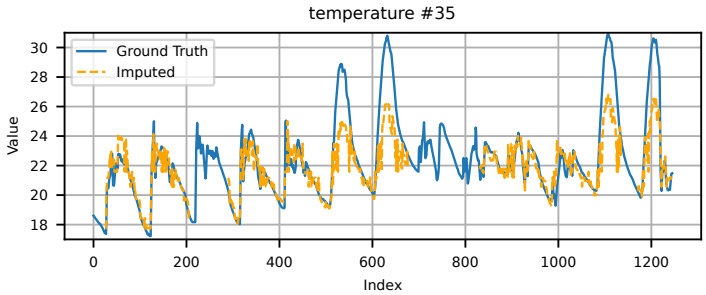
(a)



(b)



(c)



(d)

Figure 5.11: Comparison Between the Real and Imputed Values for Four Variables of the Intel Lab Dataset at 50% Missing Rate under the Consecutive Error Model.

quality. In small, highly correlated sensor networks, rule-mining methods can extract meaningful rules that are particularly effective for imputation. Our results show that GARDA surpasses traditional approaches since it first performs a partial imputation using the most reliable extracted rules, generating values that are more accurate than those produced by traditional methods alone, and then, if there are still missing values, traditional techniques can operate under more favorable conditions. This mechanism is crucial for understanding the performance gains over competing approaches. In such scenarios, GARDA can even achieve results that outperform neural network. However, this advantage decreases at low missing rates, where traditional techniques already perform well and often surpass rule-mining approaches in filling most of the gaps.

However, our method has shown that, in the context of the consecutive error model with highly correlated sensors, when there are many consecutive gaps (for instance, when sensors malfunction and do not produce values for long periods of time), it performs better than other traditional methods, and even significantly better when the missing rate is high. This is also due to the difficulty traditional methods have in performing well under such conditions. As shown in Figure 5.11, which compares the true values with the ones reconstructed by GARDA for four example sensors at a 50% missing rate, the trend of the original values is accurately captured, with only small fluctuations around the true values. A noticeable deviation between the curves can be observed only in presence of atypical spikes in the data, as in the case for temperature sensor 35. Therefore, the usefulness of GARDA in this context is even greater, as it is able to support traditional methods more effectively.

Naturally, such advantages emerge only when the GrC association-rule–based imputation process identifies and retains rules with sufficiently high support and confidence. When correlations are lower, fewer effective rules can be extracted, and most of the imputation is then performed by the traditional method (in our case, linear regression). If the process begins with a dataset containing poorly imputed values, the traditional method may also struggle to accurately reconstruct the remaining ones, leading to lower overall performance compared with other approaches. However, the simplicity of the model and its strong effectiveness on small, highly correlated datasets remain particularly valuable in IoT and edge-computing contexts, where computational and memory resources are limited. In such scenarios, the approach proves especially suitable.

5.4 Summary

In this chapter, we detailed the sensing stage of the proposed CPSS, showing how heterogeneous data are acquired and transformed into reliable inputs for the subsequent perception and comprehension modules.

We first described the acquisition phase, which combines visual sensing and contextual sensing. Visual data are collected through monthly UAV surveys performed with a DJI Matrice 300 RTK equipped with a Zenmuse L1 RGB camera and high-precision IMU/RTK georeferencing. Flight altitude and mission planning are defined to balance coverage and detail, enabling systematic monitoring of the Pompeii archaeological site. Contextual data are acquired via external sources: weather variables (temperature, humidity, rainfall, wind) are retrieved through the Open-Meteo API, while visitor-flow information is obtained from publicly available data from the Pompeii administration. In addition, we introduced the cumulative rainfall indicator R^c as a compact representation of rainfall persistence over time, which is particularly useful for interpreting vulnerability conditions beyond instantaneous precipitation values.

We then distinguished preprocessing into offline and online sensing operations. Offline sensing focuses on producing georeferenced, analysis-ready products from raw UAV imagery. Using Agisoft Metashape, experts align images, refine georeferencing with control points, generate dense point clouds, and export orthomosaics as GeoTIFFs, typically on the order of several GB. Subsequently, QGIS is used to build vector layers delimiting areas of interest, for example roofs, and to perform polygonal cropping; in the final prototype, the workflow is exemplified by extracting Regio I for integration into the dashboard described later.

Online sensing addresses data quality issues that affect both contextual streams and, more broadly, sensor-network measurements, where missingness and low-quality records are common. To tackle this problem, we proposed GARDA, Granular Association-rule-based Data Imputation Approach, a lightweight hybrid method specifically aimed at small, highly correlated IoT datasets. GARDA combines an automatic granulation stage that discretizes continuous values into interval-based information granules, supporting Equal-Width, K-Means, and KDE strategies, a granular association-rule imputation stage based on robust support and confidence definitions for incomplete data, which imputes only when rules are sufficiently reliable, and a traditional imputation stage, Bayesian ridge regression in our implementation, that completes the remaining gaps and benefits from the partial high-confidence reconstruction achieved by the rule-based phase.

Finally, we validated GARDA through an extensive experimental protocol based on artificial missingness injection. We evaluated both MCAR and

consecutive-missing patterns on the Intel Lab Dataset, temperature readings aggregated in 15-minute slots for a set of highly correlated sensors, and compared against state-of-the-art neural imputation methods, SAITS and M-RNN, on two configurations of the Beijing Multi-Site Air-Quality dataset, a high-correlation multi-station setting and a reduced moderate-correlation setting. Results show that GARDA is particularly effective at moderate-to-high missing rates and under consecutive outages, where it consistently outperforms traditional baselines and avoids the degradation observed when rule-based methods are forced to impute with low-confidence rules. In the high-correlation Beijing setting, GARDA becomes competitive with, and for higher missing rates, superior to neural approaches, while in low-correlation conditions SAITS remains superior, highlighting that GARDA's main advantage emerges when meaningful cross-sensor regularities can be mined. Across experiments, K-Means granulation provided the most robust trade-off between rule reliability, support and confidence, and reconstruction precision.

Overall, the chapter clarifies how the CPSS sensing pipeline couples expert-driven offline products, orthophotos and GIS layers, with automated online data-quality mechanisms, imputation, producing dependable inputs for downstream situation awareness. Future work will explore richer granulation models, for example fuzzy and rough granules, and alternative reconstruction schemes beyond simple interval averaging, as well as additional validation in other domains and operational deployments.

Chapter 6

Perception and Comprehension of Maintenance Issues

"Now is no time to think of what you do not have. Think of what you can do with what there is."

— Ernest Hemingway, *The Old Man and the Sea*

The perception and comprehension phases, realized through the implementation of the modules in Figure 6.1, are necessary to analyse sensor and contextual data in order to identify the elements of the environment and understand their meaning.

The perception phase operates on the data sensed through physical and virtual sensors. On its own, this raw data is not yet sufficient to be interpreted and used for decision-making. It is first transformed into information that is directly interpretable and understandable for a human or an artificial agent. Based on prior knowledge and perceived data, it can then make a decision, also taking into account how the situation is expected to evolve. This last aspect will be analysed in the Chapter 7.

In our case, this process consists in interpreting aerial orthophotos obtained from surveys of the archaeological site, together with contextual data related to weather conditions and visitors. It begins with a low-level perception, in which the fundamental features needed to analyse the situation are extracted. During a subsequent, high-level perception phase, a selection process is carried out in which only the most relevant features are retained. These features are then used to feed an AI technique that can analyse them to detect potential issues on the site and understand the overall situation, also combining the detected issues with contextual information. The goal is to assess the vulnerability of different

areas in order to support those responsible for monitoring the site in deciding which areas require maintenance.

In our case, the perception and comprehension phase focuses on the automatic identification of maintenance issues based on the orthophotos of the archaeological site of Pompeii. This task can be approached in different ways, depending on the types of maintenance issues we aim to detect. As mentioned above, in an outdoor site such as Pompeii, we are interested in identifying weedy vegetation, damaged conduits on flat roofs, and broken tiles on sloping roofs. However, this work is carried out in a complex context, namely a vast ruined urban environment, where it is difficult even to distinguish between damaged structures that need repair and ruins that must be preserved in their current state.

The identification of these issues is addressed in this chapter, and the main objectives of the related modules have evolved over time, depending on the intermediate results obtained and the critical issues identified in the proposed techniques. The criticalities that emerged in the first methodology led us to define a second one, with a revised approach that increased efficiency on the issue of greatest interest to park experts and removed most of the difficulties encountered with the first method.

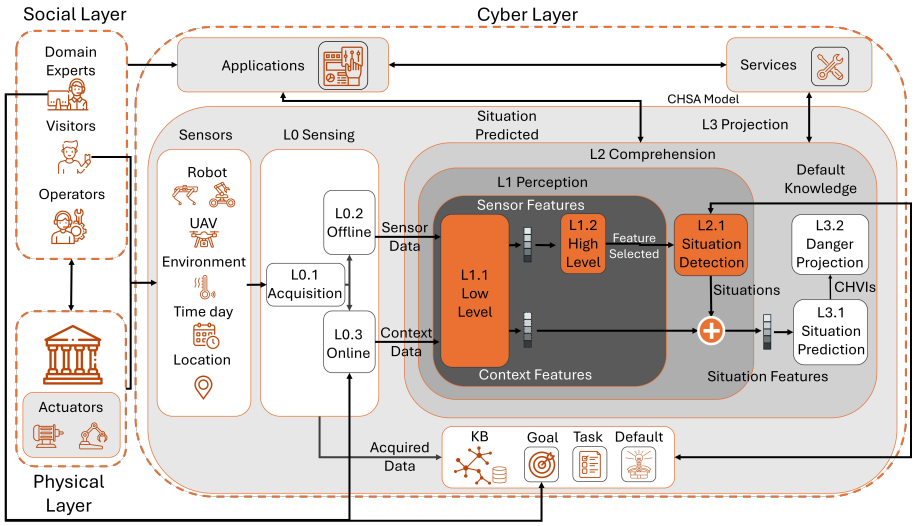


Figure 6.1: SA-CPSS with Perception and Comprehension Details

In particular, the two proposed methodologies are as follows:

1. The first solution, called **FAUNO**, is a prototype tool based on YOLO for the detection of maintenance issues on the roofs of structures in archaeological sites, starting from aerial orthophotos. FAUNO is designed

to enable fast identification of the location and classes of maintenance issues in these images. This approach assumes that the positions of the structures are unknown and operates by dividing the entire orthophoto into multiple sub images, which are then processed. In each sub image, the tool identifies the bounding box corresponding to a detected problem and the class to which it belongs. At this stage, four classes are considered: weedy vegetation, damaged conduits, damaged structures, and broken tiles. In the section 6.2 dedicated to FAUNO, the advantages and disadvantages of this approach are described, and the difficulties encountered guide the definition of a second approach focused only on roof issue classification.

2. The second solution, called **HydraML**, proposes a sustainable Artificial Intelligence method based on Fuzzy Transform. The approach assumes prior knowledge of the position of the structures, since in the field of CH, and especially in archaeological sites, the configuration of the park and its structures changes very slowly. A mapping phase of the structures, as explained in Chapter 5, can therefore be standardised and carried out quickly, avoiding an unnecessary segmentation and detection phase that would add complexity and potential sources of error, considering that the issues of greatest interest to operators concern roofs. It should be remembered that we work in an SA context, where the processes of perception, comprehension, and projection are iterative and incremental. Given the generality of our CPSS model, nothing prevents us from implementing perception and comprehension modules aimed at automatically segmenting roofs, which can then be validated in a subsequent offline sensing phase by expert users, who correct and refine the results. In this way it is possible to semi automatically determine the position of the roofs, which can then be updated over time. Once the position of the roofs is known, HydraML is applied directly to them. The method combines deep learning and non deep Automated Machine Learning models for the classification of maintenance issues affecting roofs at archaeological sites. The combination of deep and non deep models reduces training time and energy consumption compared with a purely deep learning approach. HydraML was evaluated using 10-fold-cross-validation on two real world CH datasets concerning issues on tiled and flat roofs of the archaeological site of Pompeii.

Based on these considerations and what will be discussed below, HydraML will be the method integrated into the final implementation of the CPSS in Chapter 7. The FAUNO approach is presented nonetheless, as it does not require segmentation, although this comes at the cost of lower classification

performance. Moreover, it can also identify weedy vegetation located away from building roofs. FAUNO could be particularly useful in the detection of such weedy vegetation if thermal or infrared imagery were available.

Parts of this chapter have been previously published in:

- Francesco Colace, Massimo De Santo, Rosario Gaeta, Rocco Loffredo. *Detection of maintenance issues from UAV images of archaeological sites: A yolo-based tool.* Multimedia Tools and Application. 2025. Volume 84. https://doi.org/10.1007/s11042-025-20773-7.
- Massimo De Santo, Francesco Flammini, Rosario Gaeta, Hangli Ge and Zia Ur Rehman. *An Automl Approach for the Efficient Classification of Damaged Roofs Using Fuzzy-Transform*. IEEE ICHMS. 2025. 10.1109/ICHMS65439.2025.11154372.

6.1 Background and Related Works

In this section, we analyse and discuss the literature on the detection of maintenance issues not only in the cultural heritage domain, but also in other related fields, in particular damaged tiles and conduits in urban areas, weedy vegetation, and damaged structures. We also examine the most recent and effective classification techniques used for damaged roofs, assuming that the location of the structures is already known. In addition, we address the topics of AutoML, that is paradigms for the automatic training of classification models, which are very useful in various domains and are employed in our HydraML model, and the issue of sustainability, which has gained increasing importance in recent years, as already discussed in Chapter 1.

6.1.1 Detection of Maintenance Issues

In recent years, advances in AI have led to the development of several methods and techniques for object detection. Efficiency, which is the ability of a system to perform a task with minimal resources such as time, memory, or data, has also become an important factor, with the spread of approaches based on Efficient backbones, such as EfficientDet [224], an optimisation of EfficientNet [225], and others based on YOLOV4 [226] and YOLOV5 [227].

In the following, we focus on papers that address the detection of each type

of issue we aim to identify in images.

6.1.1.1 Weedy Vegetation

One of the main maintenance issues affecting ruins within archaeological sites is the presence of weedy vegetation on the structures, which can compromise their stability. These issues must be addressed promptly and efficiently in order to prevent irreversible damage to historic buildings. Alien plant invasions are discussed in several domains: in [228], the focus is on the rapid identification and detection of alien plants, since once the invasion process is too advanced, these species become very difficult to eradicate. The proposed method relies on UAV imagery and evaluates different object-based and pixel-based detection and classification techniques. Among the main challenges are radiometric inconsistencies caused by unstable illumination, as well as errors in areas where vegetation growth is highly variable.

The work in [229] concentrates on detecting green spaces, trees, and plants in urban environments. The authors aim at identifying green regions for smart city applications. Their algorithm employs a UAV for real-time data acquisition, followed by GLCM and LBP feature extraction, and then applies a logistic regression model combined with boundary extraction techniques. The reported test accuracy is 90.32%.

In [230], the objective is to eradicate Solanum rostratum Dunal, an invasive alien plant that harms the ecosystems where it spreads and proliferates. The study presents a novel deep learning segmentation approach for detecting alien plant invasions, specifically targeting this species. UAVs are used to capture the images, which are then processed with a U-Net-based neural network called DeepSolanum-Net. The recognition precision for alien plants reaches 89.95%, with a recall rate of 90.3%.

From the literature we examined, no methods were found that are explicitly designed to detect weedy vegetation affecting buildings and man-made structures; instead, existing approaches concentrate on alien plant detection in agricultural contexts or vegetation analysis in urban areas. When weedy vegetation is specifically discussed, the emphasis is mainly on monitoring systems and platforms rather than on automatic detection using AI-based techniques.

6.1.1.2 Damaged Conduits

Another set of maintenance issues concerns the flat roofs of on-site structures. These roofs are protected by conduits to prevent water infiltration and subsequent damage to the roof itself, to the stored objects, and to the internal structural components. In the context of identifying damaged flat roofs, [231] presented

a DL system that performs damage segmentation on flat roofs using aerial imagery. The types of damage are classified as *hollows*, *swelling*, *folds*, *patches*, and *breaks*. The dataset employed by the authors consists of 6400 images. They compare three segmentation approaches: U-NetMCT, DeepLabV3+, and HRNet+OCR. The state-of-the-art multi-scale HRNet+OCR model achieves the best performance, with a mean IoU of 0.44. The fastest models are U-NetMCT and DeeplabV3+, although they obtain lower quality, with mean IoU in the range 0.33–0.37. The category "patches" is reported as the most challenging, partly because it is underrepresented in the dataset.

The study in [232] proposes a method for detecting and performing semantic segmentation of roofs from aerial images acquired after disasters. The authors use the Volan2019 dataset, which includes 16 videos and 875 fully annotated images, further expanded through data augmentation to maintain class balance. They test two CNN architectures for semantic segmentation, based on mask-RCNN and PSPNet. The best mask-RCNN model produces pixel level segmentations with a mean average precision (mAP) of 51.54%, whereas the best PSPNet models achieve a mean intersection over union (mIoU) of up to 32.17% and an accuracy of 77.01% on bulk objects. From these works, we can state that the literature in this area mainly focuses on semantic segmentation methods to detect damage on flat roofs, although relatively few studies are available. The second paper is in fact primarily oriented toward identifying different categories of roof damage related to disaster events such as hurricanes. The first paper uses a dataset entirely dedicated to flat roof deterioration, with samples photographed from short distances, while our review did not reveal any work specifically targeting the detection of deteriorated or damaged conduits from UAV based surveys.

6.1.1.3 Broken Tiles

The broken tiles class refers to roofs whose covering elements are fractured or otherwise damaged. This represents a critical maintenance issue because, even though the tiles used in archaeological sites are often not original to the historical period of the site, they still form the protective covering for many CH structures and the artworks they contain. Resolving this issue is essential, since any breakage in the roof covering can create openings through which rainwater may penetrate and damage frescoes and other cultural assets.

The problem of damaged roofs and tiles in a CH context is only rarely discussed in the literature and is more often examined in studies that focus on detecting broken roofs after disaster events, which is a closely related task. A study specifically concerning archaeological sites was conducted by Qiu et al. [98], who proposed a method based on the YOLOv8 object detector [160] to

identify roof damage in the cities of Xiamen and Quanzhou in Fujian Province. Using high resolution orthophotos acquired with a UAV, a YOLOv8 detector was applied to locate damaged roofs. The approach relies on two YOLOv8 models arranged in cascade, the first used to detect buildings and the second to identify damage on their roofs. The first stage achieves a mAP of 97.2%, while the second stage reaches 89.4%.

In [233], the authors work with remote sensing images of earthquake damage. They propose a neural network based on YOLOv5s, a Vision Transformer, and a Bi Directional Feature Pyramid Network, originally developed for detecting rural houses. The Vision Transformer is integrated into the feature extraction component of the network, and the Bi Directional Feature Pyramid Network is also employed. The method is evaluated on data from the two thousand twenty one Yangbi earthquake. The proposed model attains an average precision that is higher by 9.31% and 1.23% compared with YOLOv3 and YOLOv5s respectively, and it is 2.96 times faster than YOLOv3. The authors note that the model still needs further validation regarding its adaptability to multiple data sources and to damaged buildings at different resolutions in both rural and urban areas, in order to assess its applicability over large regions with significant resolution differences and diverse building types. The work [234] introduces and evaluates several CNNs for object detection using aerial images of disaster scenarios. The models are trained on the Volan2018 dataset to identify post disaster effects, and the authors also examine the influence of camera altitude. Eight CNN models based on the YOLO framework are tested, starting from weights pretrained on COCO or VOC and then retrained on Volan2018. The best results reach 80.69% mean average precision for high altitude imagery and 74.48% for low altitude imagery. The experiments show that models trained and tested on videos acquired at similar altitudes perform better than models trained and tested on videos captured at different altitudes.

From this analysis, we can conclude that research on broken tiles mainly concentrates on detecting damage caused by events such as hurricanes or earthquakes. In this context, the literature has not extensively examined the use of detection methods over large areas with heterogeneous building types and varying image resolutions, and most approaches rely on drone imagery collected at relatively low altitudes.

6.1.1.4 Damaged Structures

The damaged structure class includes buildings and constructions within the site that show significant deterioration and are at risk of collapse due to different causes. Similar collapse patterns can be observed in houses and structures affected by disaster events, which makes post-disaster studies highly relevant

to this context. To identify this class, we therefore refer to works that employ deep learning models for detecting collapses after disasters.

In [235], the authors propose a neural network based on YOLOv5, named DB-YOLOv5, for real-time detection of damaged buildings using UAV aerial imagery. They enhance the detector's classification capability by integrating a feature fusion module (BDSCAM). They also construct a custom dataset with images from the Beichuan and Ludian regions. The model achieves 96.36% precision and 91.38% recall for the Beichuan area, and 93.32% precision and 97.61% recall for Ludian. The authors also emphasize that the approach is intended to be applicable across different scales and data sources.

The study in [236] focuses on methods that support the recovery phase following disasters that damage buildings and cause collapses. The authors apply neural networks to Gray-Level Co-occurrence Matrix (GLCM) texture features extracted from orthophotos generated from UAV imagery. The combination of changes in GLCM texture features with an artificial neural network yields an accuracy close to 92%. However, they note that the model struggles to correctly identify structural changes associated with complete or partial building collapses.

In [237], the authors present a decision support system called (EUDSS). The method uses two backbones with shared weights to process pre-disaster and post-disaster images, and relies on features from both time points as input to a Fourier attention module that produces Fourier features. Pre-disaster images with building segmentations are then combined with a damage mask to obtain the final damage assessment. The proposed system offers the fastest inference speed while maintaining competitive accuracy compared with other approaches.

Overall, the literature provides several examples of techniques for detecting structural damage and collapses from aerial imagery, particularly in post-disaster scenarios. The reviewed works show a clear interest in improving these methods so that they can be applied more generally to large, diverse areas with heterogeneous building characteristics.

6.1.2 Classification of Damaged Roofs

In this section, we focus on reviewing the literature related to classification systems for damaged roofs. Since, to the best of our knowledge, only a few studies address roof damage classification in the cultural heritage domain, we mainly refer to work on roof damage classification after natural disasters, which is a very similar problem to the one considered here.

Cao et al. [238] apply an image classification approach to satellite imagery

acquired after Hurricane Harvey. The authors design and train a convolutional neural network and compare it with an existing neural model, achieving 97% accuracy on the test set for their convolutional neural network. Wang et al. [239] and Ishraq et al. [240] propose neural network architectures based on VGG16 to classify damaged roofs after Hurricane Harvey. Their approaches obtain test accuracy values of about 0.96% and 0.98%, respectively. Boge et al. [241] propose a method to localize and classify roof damage in aerial imagery using the concept of superpixels for damage localization. They compare a random forest classifier with two types of convolutional neural network classifiers based on DeepLabV3+ and SegNet. The damage level is determined from the ratio between damaged and undamaged roof areas. The model based on DeepLabV3+ achieves the best accuracy, reaching 98% on both datasets considered. Kumari et al. [242] present a hybrid approach that combines DenseNet201 and a Support Vector Machine, applying Principal Component Analysis as a dimensionality reduction step for the extracted features. Their method is evaluated on the Hurricane Harvey dataset and attains a promising accuracy of 97.10%.

Ramesh et al. [243] investigate the effectiveness of the Vision Transformer architecture for the classification of satellite image datasets. In their experiments, the Vision Transformer outperforms convolutional neural networks, achieving an accuracy of 88.49. Zhang et al. [244] use an ensemble of three neural networks and show that this ensemble surpasses each individual network, with test accuracies of 98%. Zarski et al. [245] present a new damage assessment method that employs an original multi step feature fusion network for the classification of building damage using large scale pre disaster and post disaster satellite images. They introduce a Fusion Module that allows any convolutional neural network to process image pairs. With this Fusion Module, all tested models achieve higher accuracy and F1 scores. Hong et al. [246] propose EBDC Net, which consists of a feature extraction encoder module and a damage classification module. The overall classification accuracy on a dataset composed of post earthquake images from three different locations is 94.44%, 85.53%, and 77.49% when two, three, or four damage classes are considered, respectively.

6.1.3 Sustainable AI

Sustainable AI can be understood as the design and development of AI methods that seek to minimize negative environmental, social, and economic effects while supporting long term human and ecological well-being [247]. This idea is attracting increasing interest as the scale and adoption of AI technologies continue to grow, together with their environmental footprint. In particular, the

training and inference stages of large and deep AI models often require significant computational resources, raising concerns about their resource intensity and long term viability. As a result, the research community is placing more emphasis on limiting the environmental impact of these approaches [248].

A key topic in current research is the energy efficiency of AI models. This involves the development of new learning algorithms, the optimization of architectures so that they perform well on low power or specialized hardware, and advances in software hardware co design. There is also a growing interest in measuring the environmental footprint of AI systems and in reducing training time while preserving strong performance.

At the same time, researchers are examining more carefully the role of data in sustainable AI. Instead of simply enlarging datasets, which can be both resource heavy and ethically problematic, there is a move toward improving data quality and data governance [249].

In parallel, the notion of sustainability is being broadened beyond ecological aspects alone. Increasingly, scholars and practitioners stress the social and ethical dimensions of sustainable AI, for example by seeking to ensure that AI systems do not deepen existing inequalities or introduce new forms of harm.

These aspects are therefore fully aligned with the 2030 Agenda for Sustainable Development, discussed at the beginning of Chapter 1, and our research aims to support the seven SDGs mentioned there.

6.1.4 AutoML

AutoML focuses on automating the selection of the best classifier or regression model to use [250]. Popular approaches include methods based on reinforcement learning and evolutionary algorithms. Techniques based on Genetic Programming [251] or on multi stage stacked ensembling are often used in AutoML to search for the best models among traditional methods, while Neural Architecture Search (NAS) [252] is commonly employed to find the best neural network in a given search space.

In AutoML, the automation of the machine learning pipeline is achieved by optimising both algorithm selection and hyperparameter tuning. The combination of algorithm selection and hyperparameter optimisation (CASH) has been introduced as a way to address both tasks simultaneously [253]. Unlike traditional approaches, where algorithm selection and hyperparameter optimisation are treated as two separate steps, in the CASH problem they are solved together.

The CASH problem can be formulated as the minimisation of a loss function L over a set of algorithms $\mathcal{A}$ and their hyperparameters γ. Given a dataset D,

it can be defined as:

$$(\mathbf{A}, \boldsymbol{\gamma})^* \in \arg \min_{\mathbf{A} \in \mathcal{A}, \boldsymbol{\gamma} \in \Gamma} R(P_g, \mathbf{A}, \boldsymbol{\gamma}, D) \tag{6.1}$$

where $\mathbf{A} \in \mathcal{A}$ is the chosen algorithm and $\boldsymbol{\gamma} \in \Gamma$ is the set of hyperparameters associated with that algorithm. The function $R(P_g, \mathbf{A}, \boldsymbol{\gamma}, D)$ evaluates the performance of the predictive model P_g, represented by the algorithm $\mathbf{A}$ with hyperparameters $\boldsymbol{\gamma}$, on the dataset D.

Each algorithm $A \in \mathcal{A}$ has a corresponding hyperparameter space denoted by Γ_A, and the overall configuration space for all algorithms is the Cartesian product of these individual hyperparameter spaces. Solving the CASH problem effectively enables AutoML systems to select the best algorithm and hyperparameters automatically, achieving optimal model performance without manual intervention.

6.2 FAUNO: YOLO Detection Approach

The FAUNO tool we propose is able to analyse orthophotos of archaeological sites using DL for the automatic detection of maintenance issues. It examines different types of issues present in the input orthophoto and reports them to users in a dedicated document that supports their decision-making. We analysed various YOLO-based [254] object detection models in order to select the most suitable one.

FAUNO is designed to rapidly identify the positions and classes of maintenance issues in a large archaeological site orthophoto. First, the orthophoto is divided into multiple sub images, which are then processed. In each sub image, the tool identifies the bounding box corresponding to a detected issue and the class to which it belongs. Four classes are considered: weedy vegetation, damaged conduits, damaged structures, and broken tiles.

In particular, the following subsection formally describes the approach and then presents its evaluation, including the custom dataset used, the chosen evaluation protocol, the results, and the related discussion.

6.2.1 FAUNO Approach

In Figure 6.2 the tool's architecture is shown. Given an orthophoto O of dimension $d = w \times h$ where w is the width and h is the height and $d \geq O(10^8)$ pixels, an AI model cannot process a photo of such size. We need to split O in smaller images processable by a detection module. For this reason, the batch generator module acts as a sliding window that flows over the photo and

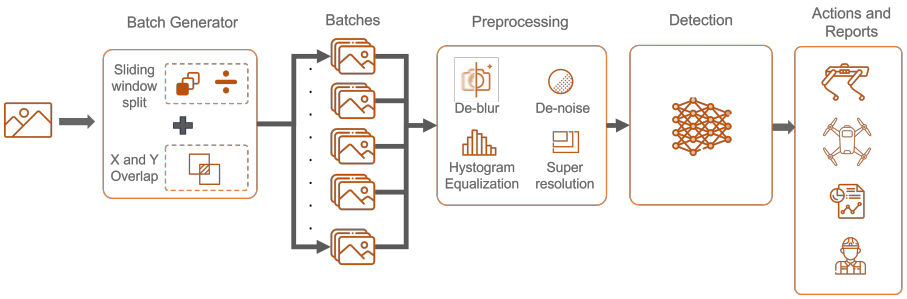


Figure 6.2: FAUNO Approach

divides it into various batches. Since typically some maintenance issue (more in general, an object) could end up partly on one photo and partly on another, it is convenient to split the photo with an overlap, that must be decided based on the issue we are dealing with and the dimension of the batches we create. Given p the percentage of overlap, s_w the width of the sub-images and s_h and the height of the sub-images, we define the set of resulting sub-images from the split as $I = \{I_1, I_2, \ldots, I_n\}$ where each sub-image I_i has dimensions s_w and s_h. We also define $o_w = [p * s_w]$ as the number of pixels for horizontal overlap and $o_h = [p * s_h]$ the number of pixels for vertical overlap. The number of sub-images generated at this point can be easily calculated:

$$n_w = \frac{(w - o_w)}{(s_w - o_w)} + 1 \tag{6.2}$$

n_w is the number of images that must be generated for the horizontal overlap;

$$n_h = \frac{(h - o_h)}{(s_h - o_h)} + 1 \tag{6.3}$$

n_h is the number of images that must be generated for the vertical overlap. The total number of sub-images n will be:

$$n = n_w \cdot n_h \tag{6.4}$$

From this each image can be defined as:

$$I_{ij} = O[x_i : x_i + s_w, y_j : y_j + s_h] \tag{6.5}$$

where $x_i = i \cdot (s_w - o_w)$, $y_j = j \cdot (s_h - o_h)$ with $0 \leq i < n_w$ and $0 \leq j < n_h$. The set I of images can be defined as $I = \{I_{ij} \,|\, 0 \leq i < n_w,\ 0 \leq j < n_h\}$. At this point, each I_{ij} can be preprocessed to increase the image quality to find maintenance issues.

The aerial images we must process may be affected by shadows, noise and blur caused by image acquisition, or low resolution. In such a context some preprocessing transformation can be used to solve or limit such problems. Given $P = \{P_1, P_2, \ldots, P_m\}$ with m number of the preprocessing function, we define $I^p_{ij} = P_m \circ P_{m-1} \circ \ldots \circ P_1(I_{ij})$ the preprocessed I_{ij}. The final pre-processed set is $I^p = \{I^p_{ij} \,|\, 0 \leq i < n_w,\, 0 \leq j < n_h\}$.

Afterwards, a detection model must be applied to localize all the possible issues in each image. Such model must convert each I^p_{ij} in a set of features $\phi^p_{ij} = \{\phi^p_{ij1}, \phi^p_{ij2}, \ldots, \phi^p_{ijk}\}$ where k is the number of resulting features. Such features must undergo a regression and classification process to find the location of the problem within the image and the class it belongs to.

In particular, we can define our model as:

$$M(\phi^p_{ij}) = \{(B^p_{iju}, C^p_{iju}, S^p_{iju}) \,|\, 0 \leq i < n_w,\, 0 \leq j < n_h, 0 \leq u < n_b\} \tag{6.6}$$

where $B_{iju} = (x_{iju1},\, y_{iju1},\, x_{iju2},\, y_{iju2})$ represents the bounding box coordinates of the i^{th} and j^{th} detected objects, with $(x_{iju1},\ y_{iju1})$ and $(x_{iju2},\, y_{iju2})$ representing the top-left and bottom-right corners of the box, C_{iju} is the predicted class label for the object and $S_{iju} \in [0, 1]$ is the confidence score associated with the detection.

Once the issues have been detected in each image, we need to determine their position in the real world, taking advantage of the fact that the images are georeferenced. For simplicity, we assume that the position of the object in the image is given by the centre of the bounding box. For any bounding box B_{iju} is possible to calculate such point as: $(x^c_{iju},\, y^c_{iju}) = (\frac{(x_{iju1}+x_{iju2})}{2},\, \frac{(y_{iju1}+y_{iju2})}{2})$ To convert such point in the image into real world coordinates we define a function:

$$G : (x^c_{iju},\, y^c_{iju}) \rightarrow (X^c_{iju},\, Y^c_{iju}) \tag{6.7}$$

Where $(X^c_{iju},\, Y^c_{iju})$ are the coordinates in the real world.

Then the found issues are used to create a report that for each of them shows the image of the issue, the position $(X^c_{iju},\, Y^c_{iju})$ of the issue, its class and confidence, in a way that the human operators can find the issues and make decisions about it. Obviously, any duplicate images must be removed, which is possible since the orthophotos are georeferenced.

6.2.2 FAUNO Detector Tool

The FAUNO approach has been implemented into a tool namely FAUNO Detector. The tool described in Figure 6.2 has been implemented using the Python programming language. As shown in Figure 6.3, the tool requires

Figure 6.3: FAUNO Design Architecture

the user to upload an orthophoto of the site and to select a detection model. The orthophoto is then opened and processed using the GDAL library, which enables splitting the orthophoto into multiple patches. The size of the patches is critical, as it influences the material on which the position and class of the issues in the site are inferred. After a grid search, the chosen split size was 1280×1280 with $o_w = o_h = 0.5$.

At this point, the patches can be preprocessed to improve their quality before starting inference. Images with too many white pixels, due to contour images or merging defects, are filtered out. This filtering is performed by imposing a threshold of 10%, so that all images with a percentage of white pixels greater than this threshold are removed. Histogram equalization and Gaussian deblurring are also applied to mitigate illumination issues and noise.

The YOLO detector is then applied to each patch. The motivation for selecting this family of models lies in its well known balance between speed and

accuracy, which enables rapid and reliable object identification [255] [256]. YOLO is designed to be very fast and suitable for real-time applications, ensuring high processing speed compared with other methods that analyse the image multiple times, such as RCNN and Fast RCNN [254]. Although our application is not used in real-time, we must analyse very large images that need to be divided into many patches. Indeed, considering typical orthophoto dimensions of about 50000×30000 pixels, a patch size of 1280×1280 and a horizontal and vertical overlap of 0.5, a general orthophoto will produce approximately 3600 patches. For this reason, we require a fast and efficient technique in order to obtain a tool that is practical and feasible to use, also when considering the time needed to split the orthophoto, reconstruct the position of the detected objects in the real world, and generate the issue report.

In addition, YOLO has an accurate global view of the image, tends to make fewer background errors, and can detect multiple objects within the same image simultaneously [255] [257]. Both of these characteristics are important and fundamental in our context, since each patch is very complex to analyse, with many elements, multiple issues to detect, and heterogeneous landscapes and backgrounds. Using such a detector, for each patch an image is generated and saved with bounding boxes surrounding each detected maintenance issue and indicating its class. A `.txt` file with the predicted labels is also produced for each image, along with a `.gpkg` file that can be used in QGIS to visualise all the identified issues.

The user can then press a button to generate a PDF report of the detected maintenance issues. In addition to general information and statistics, the report lists the issues found for each class, together with an identifier, longitude, latitude, and a map link that shows operators the precise location where each maintenance issue was detected. To achieve this, once a rectangle around a maintenance issue has been identified, the centre of the rectangle is computed. This centre is then converted from a position in the image to a position in the coordinate system in which the image is georeferenced, in our case EPSG:3004 (also known as Monte Mario / Italy Zone 2). For this work, the Rasterio library, in particular its `transform` module, was used for this operation.

However, to be displayed in Google Maps, the coordinates must be expressed in an interpretable reference system, usually EPSG:4326, also known as World Geodetic System 1984 (WGS84). If the tool detects that the coordinate system in which the image is georeferenced differs from this, it automatically performs the conversion and produces a Google Maps link, which is then included in the report. For this final conversion, the `ogr` and `osr` modules of GDAL are employed.

6.2.3 Evaluation

In this section, we evaluate our approach through a case study on the archaeological site of Pompeii. We first describe the dataset used in the experimental phase and the evaluation protocol, including the family of models investigated for our tool and the metrics adopted. We then present and discuss the results, providing both a quantitative analysis of the detection model and a qualitative analysis of the tool, with comments on its limitations and possible future developments.

6.2.3.1 Dataset

The dataset was obtained starting from six orthophotos related to the months of April, July, August, September, October, and November 2022. The creation of the dataset begins with expert staff identifying issues on the orthophotos of the archaeological site of Pompeii. Domain experts identified five different maintenance issues for each image:

- **Weedy Vegetation** that could infiltrate structures and cause damage;
- **Damaged Conduits** on flat roofs that must be repaired to prevent water infiltration;
- **Fallen Elements** from walls that could indicate an advanced state of deterioration and thus the need for intervention;
- **Water Accumulations** on floor or roofs which can generate infiltration;
- **Disconnected Elements** that describe a range of situations, including damaged structures and foreign objects on roofs and roads.

The experts have labelled each maintenance issue by placing points on the affected area using QGIS. These points have been saved into `.gpkg` files, and the coordinate system used is EPSG:3004, also known as Monte Mario / Italy Zone 2. The dataset was carefully analysed, and modifications were made to improve the training efficiency:

- Fallen elements were excluded from the analysis because of their similarity to stones in the area. In addition, the difficulty to distinguish, in orthophotos, between parts of collapsed walls and ordinary stones, especially during the labelling phase, also contributed to this choice.
- Water mounds were excluded from the analysis because of their ambiguity with respect to roofs, which made the *water mound* and *damaged*

conduit classes too similar. Except for a few rare cases, ground level water mounds were also found to be highly heterogeneous and difficult to label in a way that would be useful for effective network training.

- The last class we decided to remove is *disconnected element*, due to the very high heterogeneity it exhibits in the available dataset. This class includes unique maintenance issues, which led to instances appearing either in the training set or in the test set, but not in both.
- It was decided to replace the *disconnected element* class with two separate classes: one for broken tiles on roofs and one for damaged or collapsed structures.

In the end, the resulting dataset used to train the neural network has four classes:

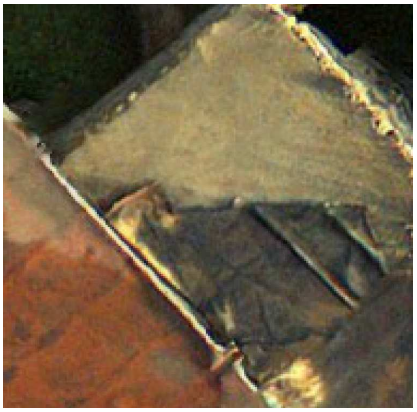
- Weedy Vegetation;
- Damaged Conduit;
- Damaged Structure;
- Broken Tile.

In the context of an object identification algorithm, it was necessary to process the orthophotos and label points into smaller orthophotos and bounding boxes, respectively, to create a processable dataset. In particular, the points that the experts indicated in the orthophoto were transformed into a bounding box of a default standard dimension. Then, using a labelling tool named *labelme* (`https://github.com/wkentaro/labelme`), the bounding boxes were manually adjusted to fit the affected areas as accurately as possible. Bounding boxes were placed around damaged conduits and broken tiles to encompass the affected roof. In contrast, bounding boxes around weedy vegetation and damaged structures were placed to encompass only the specific areas of concern.

Looking at the images in Figure 6.4, we want to specify that we labelled samples of weedy vegetation (a) as all type of weedy vegetation in proximities of walls and roofs that can damage the structures, damaged conduit (b) as the flat roof in which the conduit is considerable damaged, with humidity stains, with conduits partially removed or totally removed, broken tile (c) as roofs with partially or totally damaged tiles, damaged structures (d) severely damaged and uncovered.

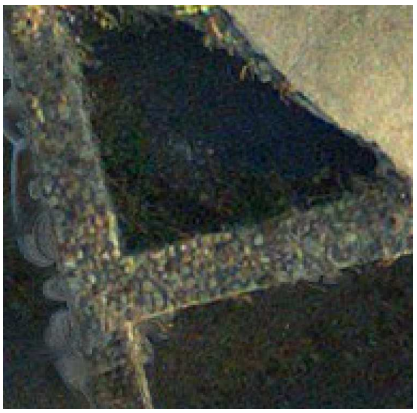
As a result, a dataset was created containing 901 images of size $1280x1280$ with a total of 2157 bounding boxes, including 1021 weedy vegetation, 711 damaged conduits, 144 damaged structures and 281 broken tiles. From the

(a) Weedy Vegetation



(b) Damaged Conduit

(c) Broken Tile



(d) Damaged Structure

Figure 6.4: Maintenance Issues Examples

dataset, as shown in Table 6.1, we randomly selected a 20% test set and perform a 5-fold cross validation to tune the hyperparameters and choose the model.

Type of Damages	**Train/Validation**	**Test**
Total Images	722	179
Weedy Vegetation	637	222
Damaged Conduit	487	119
Damaged Structure	87	31
Broken Tile	184	60

Table 6.1: Division of Images and Labels into Train/Validation, and Test Sets.

Since we do not have much data, we decided to fine tune pre-trained models on similar images and to augment the dataset in order to increase the number of examples available to our model. In particular, we applied a series of transformations to the images in the training dataset during the training epochs to avoid overfitting and to make the model more robust to variations in the characteristics and features of the images. Indeed, during acquisition, various conditions can change.

The data augmentation techniques we used were of multiple types. Colour oriented data augmentation is based on modifying the hue, saturation and value of an image to make the network more robust to different acquisition conditions. The same area photographed at different times of the day may appear very different because of lighting conditions [258]. To increase the variability of the dataset, it was also useful to apply image rotation, which in our case is +/- degree. We also used left and right flipping, applied with a probability of percent. Another type of data augmentation we adopted is mosaic, which is applied during the training phase with a probability of 30%. We also introduced Gaussian noise injection and a random crop with 20% probability.

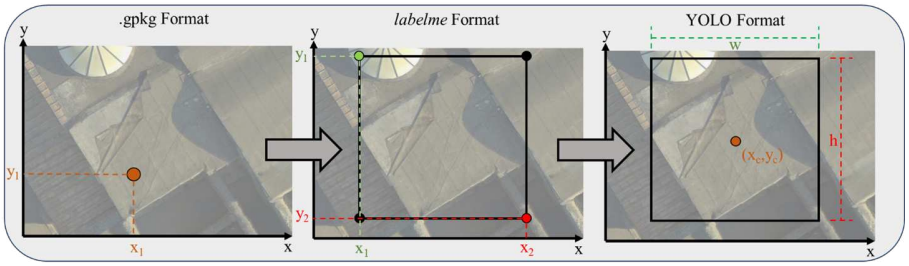


Figure 6.5: FAUNO Labelling Format Conversion

The annotation format provided by the `.gpkg` files consists essentially of points, which are converted into bounding boxes using *labelme*. In this case, the new format requires two pairs of coordinates $(x1, y1)$ and $(x2, y2)$, where $x1$ and $y1$ are the coordinates of the top left corner of the bounding box, and $x2$ and $y2$ are the coordinates of the bottom right corner.

A Python program was written to convert this format into the YOLO one. In our experiments, we investigated various versions of YOLO. In particular, we tested several architectures of YOLOV5, YOLOV7 [256], and YOLOV8, with a fine tuning of the weights trained on the COCO dataset and made available by the respective developers.

Since we work with a YOLO detector, during our experiments we focused on converting the *labelme* bounding box coordinates into a format compatible with YOLO, that is centre, height, and width, as illustrated in Figure 6.5.

In the *LabelMe* format, the bounding box is defined by two manually selected points (x_1, y_1) and (x_2, y_2) in image coordinates (top-left and bottom-right points of the bounding box). To generate labels in the YOLO format, each object is represented by five values: the class identifier and the normalized bounding-box parameters $(x_c^n, y_c^n, w_{box}^n, h_{box}^n)$, where (x_c^n, y_c^n) is the box center and (w_{box}^n, h_{box}^n) are its width and height.

To make the conversion robust to the order in which the two points are provided, we first compute the minimum and maximum coordinates:

$$x_{\min} = \min(x_1, x_2), \qquad x_{\max} = \max(x_1, x_2), \tag{6.8}$$

$$y_{\min} = \min(y_1, y_2), \qquad y_{\max} = \max(y_1, y_2). \tag{6.9}$$

The bounding-box width, height, and center in pixel coordinates are then:

$$w_{box} = x_{\max} - x_{\min}, \qquad h_{box} = y_{\max} - y_{\min}, \tag{6.10}$$

$$x_c = \frac{x_{\min} + x_{\max}}{2}, \qquad y_c = \frac{y_{\min} + y_{\max}}{2}. \tag{6.11}$$

According to the YOLO format, these quantities are normalized to the range $[0, 1]$ by the image width w_{image} and height h_{image}:

$$w_{box}^n = \frac{w_{box}}{w_{image}}, \qquad h_{box}^n = \frac{h_{box}}{h_{image}}, \tag{6.12}$$

$$x_c^n = \frac{x_c}{w_{image}}, \qquad y_c^n = \frac{y_c}{h_{image}}. \tag{6.13}$$

For each image, we created a `.txt` file containing one row per object in the form `class_id` x_c^n y_c^n w_{box}^n h_{box}^n, where the class identifier is followed by the normalized YOLO bounding-box parameters computed with the equations above.

6.2.3.2 Evaluation Protocol

In this section, we describe the experimental phase that led to the choice of the predictive model to be used, and we comment on the results, limitations, and possible future developments of the chosen approach. In particular, the machine on which the training/testing operations were executed is equipped with:

- CPU: 13th Gen Intel(R) Core(TM) i9-13900K, 3000 Mhz, 24 core, 32 logic processor;
- RAM: 64 GB;

- GPU: RTX A6000 with 48 GB of dedicated RAM.

We evaluated several models from the most recent and widely used YOLOv5, YOLOv7, and YOLOv8 families. YOLOv5 and YOLOv8 are versions developed by Ultralytics on top of the original YOLO models, whereas YOLOv7 is an evolution proposed by the original YOLO authors. The evaluation was carried out by fine tuning these models, starting from publicly available pretrained weights and adapting them to the dataset described above. For our training phase we set Adam as the optimizer, with:

- Initial Learning rate $lr0$: 0.0001
- Final OneCycleLR lrf: 0.001
- Momentum (Adam $beta1$): 0.900
- Weight Decay: 0.0005
- Warmup Epochs: 4.0
- Warmup Momentum: 0.933
- Warmup bias lr: 0.1

These final settings are the result of a grid search to optimise some of the hyperparameters mentioned above, in particular $lr0$ and the momentum. Since the final learning rate is $lr0 * lrf$, and given that each cross validation requires many hours, we decided to optimise only $lr0$. For the same reason, we did not perform an exhaustive grid search for every parameter, although we still tested several combinations, among which the one presented here turned out to be the best. All the networks we tested were pre trained on images of size 640, so we decided to resize our images to 640×640.

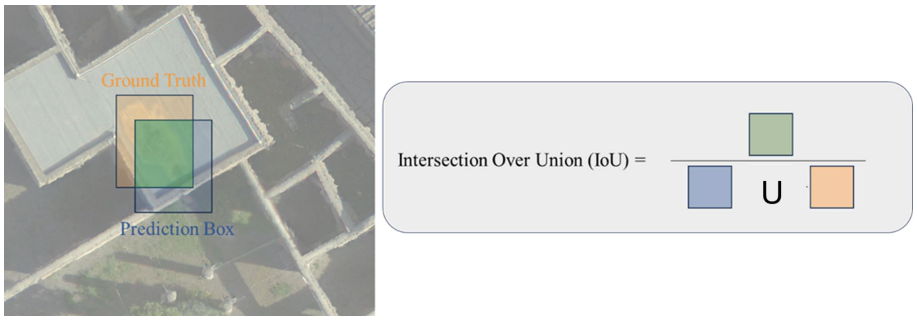


Figure 6.6: FAUNO Visualization of Intersection Over Union (IoU)

The metrics we used for our experimentation are the classic Precision (P), Recall (R), and F1-Score (F1). For the multiclass setting, we adopted their

macro-averaged versions, computed by calculating the metrics independently for each class and then averaging across all classes.

$$P = \frac{TP}{TP + FP} \tag{6.14}$$

$$R = \frac{TP}{TP + FN} \tag{6.15}$$

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{6.16}$$

where TP are True Positives, FP are False Positives, and FN are False Negatives.

The mAP50 is calculated as:

$$\text{mAP50} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i^{\text{IoU}=0.50} \tag{6.17}$$

while the mAP50-95 is the mAP averaged over IoU thresholds from 50% to 95% (at 5% steps), with AP being the Average Precision, which is the approximation of the area under the PR curve, and the Intersection over Union (IoU), which indicates the overlap between the predicted and ground-truth bounding box (Figure 6.6).

Considering such metrics, we investigated 5 dimensions in terms of layers and parameters for both YOLOV8 and YOLOV5 from nano (n) to extreme-large (x) and three versions of YOLOV7, tiny, w6 and d6. The models were trained on 300 epochs but with a scheduler which interrupt the training phase if no improvement in terms of the metrics involved.

6.2.3.3 Results

The quantitative analysis compares the performance of the various YOLO models considered. Table 6.2 shows that the models with the best overall cross validation results are YOLOv5l and YOLOv8l. In particular, YOLOv5l achieved an F1 score of 0.482, 0.427 for mAP0.5, and 0.264 for mAP0.5–0.95, while YOLOv8l obtained an F1 score of 0.481, 0.427 for mAP0.5, and 0.271 for mAP0.5–0.95. YOLOv8x also shows good performance, although with a lower F1 score compared with YOLOv5l. The results are summarised in Figure 6.7, which provides a graphical comparison of the overall performance.

Considering that YOLOv8 is about three times faster than YOLOv5, we would be inclined to choose YOLOv8 based solely on cross-validation results. However, we performed an additional comparison between the two models on the test set. In particular, Table 6.2 shows that the performance of YOLOv5l is overall better than that of YOLOv8l. YOLOv5l is able to generalise better

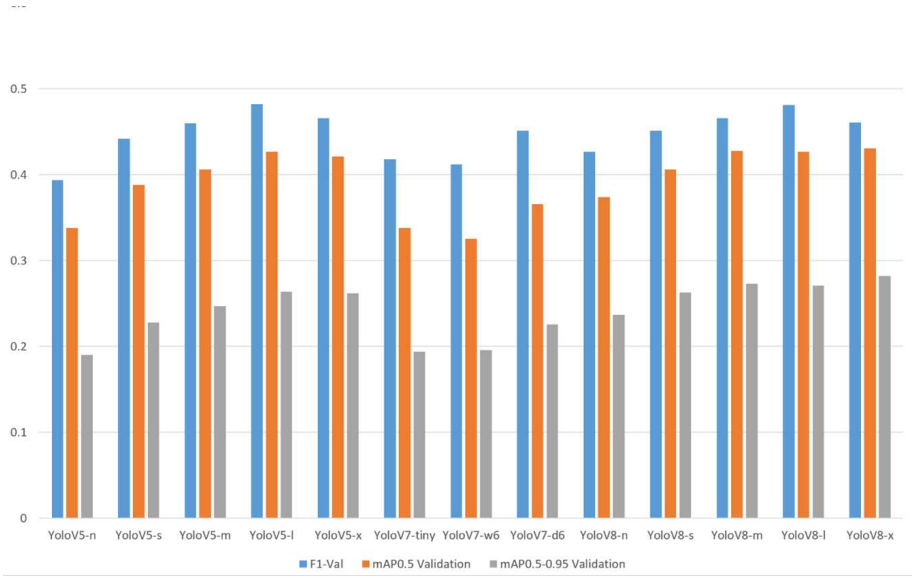


Figure 6.7: Bar Plot of $mAP_{0.5}$ Results

across classes, losing some performance on damaged conduits but gaining on broken tiles. Specifically, YOLOv5l achieves an F1-score of 0.502, a Precision of 0.564, a Recall of 0.45, a mAP50 of 0.482, and a mAP50–95 of 0.279. Although YOLOv8l has better Precision, in our case we prefer a higher Recall, since potential additional false positives can still be filtered by human operators, who can verify whether the detected issues actually exist or not.

Classes	**F1 Test**	**Precision Test**	**Recall Test**	**mAP50 Test**	**mAP50-95 Test**	**Model**
All	0.483	**0.605**	0.402	0.419	0.258	YOLOv8l
	0.502	0.564	**0.453**	**0.482**	**0.279**	YOLOv5l
Weedy Vegetation	0.345	0.611	0.240	0.284	0.138	YOLOv8l
	0.328	0.508	0.242	0.375	0.182	YOLOv5l
Damaged Conduit	0.657	0.656	0.658	0.702	0.477	YOLOv8l
	0.618	0.591	0.647	0.587	0.415	YOLOv5l
Damaged Structure	0.404	0.578	0.310	0.319	0.144	YOLOv8l
	0.412	0.712	0.290	0.466	0.152	YOLOv5l
Broken Tile	0.471	0.573	0.400	0.371	0.273	YOLOv8l
	0.522	0.444	0.633	0.500	0.367	YOLOv5l

Table 6.2: Comparison between Test Set Results of YOLOV5-l and YOLOV8-l

In Table 6.3 we report a summary of precision, recall, and F1 score for both cross-validation and the test set. In particular, the model appears to be weak

on weedy vegetation and damaged structures, while it shows higher and more stable performance on roofs.

Considering the computational efficiency of the approach, the YOLOv5l model took about 13 seconds per epoch, for a total of 188 minutes to complete 967 epochs. The FPS of the model is 54.945. For an orthophoto of about 10 GB and a split into 1280×1280 patches, it takes about seven minutes to analyse the entire orthophoto and three minutes to generate the report.

Classes	F1-Val	Precision Val	Recall Val	F1-Test	Precision Test	Recall Test
All	0.482	0.531	0.449	0.502	0.564	0.453
Weedy Vegetation	0.339	0.467	0.269	0.328	0.508	0.242
Damaged Conduit	0.671	0.679	0.671	0.618	0.591	0.647
Damaged Structure	0.461	0.542	0.413	0.412	0.712	0.290
Broken Tile	0.431	0.434	0.442	0.522	0.444	0.633

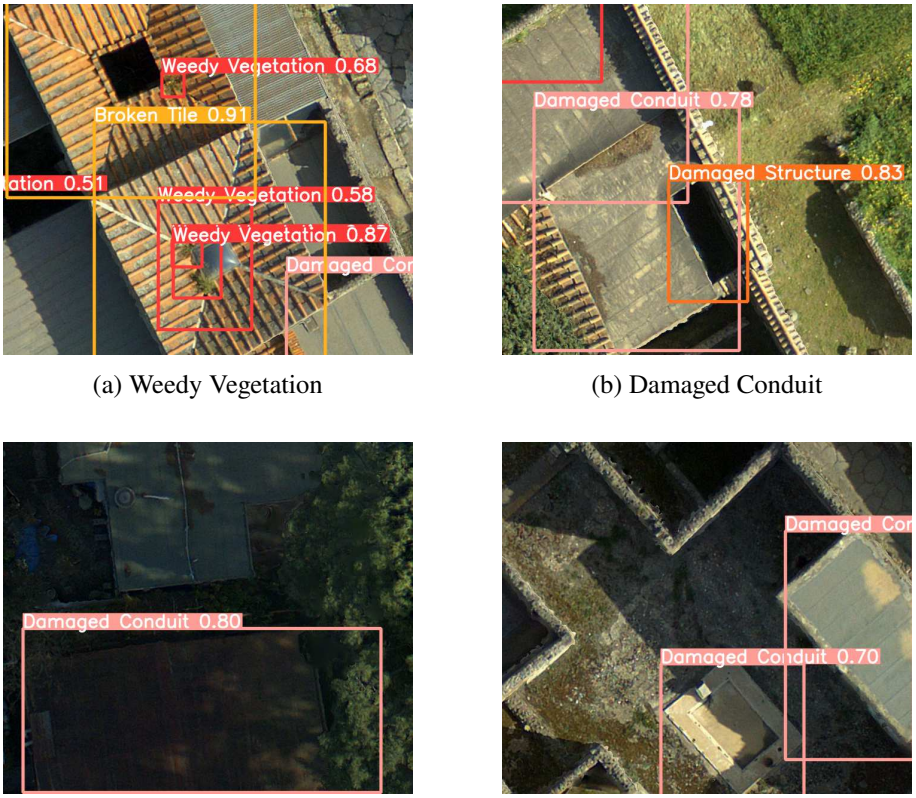
Table 6.3: F1-Score, Precision and Recall in Cross-Validation and Test on YOLOV5-l

6.2.3.4 Discussion

The description of the FAUNO approach and its results highlight both the strengths and the weaknesses of the proposed solution.

From the quantitative analysis, we observed that the approach has several limitations, especially for some classes, and although the results are useful as a proof of concept, they need to be improved before being adopted in a real system. One of the main difficulties arises from the custom dataset, which comes from an archaeological site where the environmental characteristics are very varied and complex, ranging from natural to urban and strictly historical ruined scenarios. In a context like this, even for a human operator it can be difficult to identify the relevant issues, including during the labeling process.

The identification of damaged conduit and broken tiles produced more encouraging results, especially for damaged conduits, which maintains a good recall despite the complexity of the dataset in both validation and test. On the other hand, weedy vegetation and damaged structures are more challenging to classify. This is due to the strong variability in shapes and in the types of maintenance issues involved in both cases. For weedy vegetation, the problem can range from large, rooted structures on ruined walls to small clumps of grass on walls and roofs. Damaged structures are also highly variable and, in an environment already characterised by ruins, can easily confuse predictive models during training.



(a) Weedy Vegetation

(b) Damaged Conduit

(c) Broken Tile

(d) Damaged Structure

Figure 6.8: Examples of Detection of Maintenance Issues in Images

It should also be noted that the results are influenced by the difficulty of constructing the dataset in this domain, where the labeling of maintenance issues is often borderline and challenging, especially for the weedy vegetation and damaged structure classes. A significant part of the errors we observed can be attributed to the limited diversity of the dataset, which restricts the model's ability to confidently distinguish between certain types of structures and conditions.

In Figure 6.8 (a), the model mistakenly identifies broken tiles in cases where the roof exhibits unusual features such as vegetation, central structural gaps, and small areas of deterioration and mold, which lead the system to consider the roof damaged. In this case, the misclassification is not critical, since the roof is indeed deteriorated, although with low severity. Figure 6.8 (b) shows a case where a gap in the roof is classified as *Damaged Structure* because the model is not sufficiently familiar with this specific architectural feature and confuses a roof opening with a structural collapse. A larger number of examples of this

type of architecture could help avoid such mistakes.

In Figure 6.8 (c), shadows on conduits are misinterpreted as structural damage. Shadows limit visibility and introduce on screen elements that can resemble moisture stains and deterioration. Our approach has proved not very robust in these situations, given the difficulty of distinguishing maintenance issues from shadows in aerial images. Figure 6.8 (d) presents a case where the model confuses pools with roofs with damaged conduit. The top down view typical of orthophotos makes it difficult to perceive depth, and the textures of hollows in the ground can resemble conduits or roofs. When such hollows show damage or simple flooding and stains, they can easily be interpreted as roofs.

Discussions with domain experts following these experiments prompted us to reconsider the problem and to make different choices regarding both the overall approach and the techniques adopted.

6.3 HydraML: a Hybrid Damaged Roofs AutoML Classifier for Cultural Heritage

The FAUNO approach described above has shown limitations in identifying maintenance issues. The main criticalities emerge from the mismatch between the quality of the dataset and the complexity of the problem. The orthophotos are highly complex, containing many on screen elements, and are affected by shadows and varying brightness, as well as by the presence of very different types of environments that must be analysed together. The results showed that detecting damaged structures is particularly challenging in an urban environment of ruins, where distinguishing between expected and unexpected collapses is complex. The system also struggles with weedy vegetation, due to the great variability in shapes, locations, and supporting structures. Identifying both the position and the class of issues in such a difficult context, combined with the limited dataset available, proved to be a very demanding task. This led us, in agreement with the experts, to consider simplifying the problem.

We decided to focus on the identification of roof issues, as these are the most critical. Damaged roofs can lead to structural collapse or to damage of artefacts and interior frescoes due to water infiltration and weather effects. Rather than attempting to automatically identify all the structures in the site using AI techniques and thereby introducing additional points of failure, we can exploit the fact that, in an archaeological site, the structures to be monitored are already known and the discovery of new ones is rare and slow, so updates can be made quickly and with limited effort. Given the availability of georeferenced

orthophotos such as ours, it is possible to introduce an initial configuration phase in which the structures to be monitored are defined and then extracted in all new orthophotos. This setup phase could also be automated using detection or segmentation techniques, but in our domain this may be unnecessary and even counterproductive if performance is not sufficiently high. Moreover, the previous technique required a large amount of time and data for training, with a non negligible impact on the environment.

For these reasons, we propose a sustainable AI solution named HydraML: Hybrid Damaged Roofs Automated Machine Learning Classifier for Cultural Heritage, based on Fuzzy Transform. The method combines deep learning and non deep Automated Machine Learning models for the classification of maintenance issues affecting roofs in archaeological sites. The integration of non deep techniques is intended to reduce the training time required by the deep feature extraction model and consequently lower energy consumption. The aim of this proposal is to improve performance and make the system reliable in identifying roof maintenance issues, while providing a lightweight and flexible solution that is also environmentally sustainable.

6.3.1 Methodology

The perception and comprehension phases shown in Figure 4.1 are implemented through the novel HydraML.

Our approach uses aerial images of roofs acquired by UAVs. One or more drones survey the area, capture images of the buildings, compress them to reduce communication energy consumption and mitigate noise, and send them to the cloud to be processed by the hybrid approach. The images are transmitted in compressed form to a central server, which speeds up transmission, reduces energy usage, and helps clean residual noise in the images.

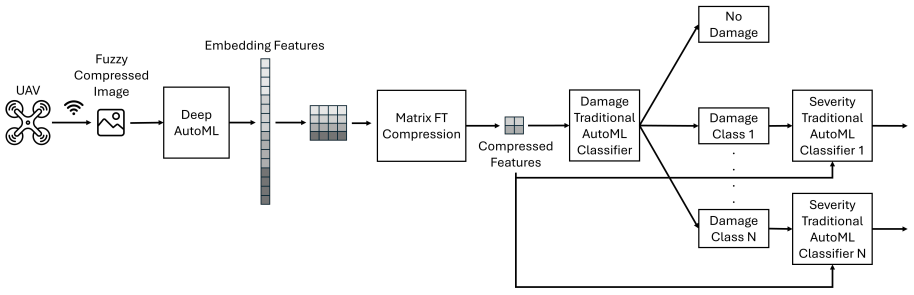


Figure 6.9: HydraML: Hybrid Damaged Roofs AutoML Classifier for Cultural Heritage

Once an image is acquired, it is provided to a neural network selected during the training phase through a Deep AutoML approach, which chooses the best combination of layers for the network. The neural network is then used to extract embeddings from the image, and the resulting one dimensional feature vector is reshaped into a two dimensional matrix and passed to a Fuzzy Transform based compression function, which reduces the number of features and the dimensionality of the classification problem. These compressed features are then used to train an AutoML approach that selects the best traditional classifier. This classifier decides whether a roof is damaged or not and, in the case of damage or other issues, a second AutoML approach in cascade is used to classify the severity of the damage or issue.

In the following, the AutoML and Compression approaches will be discussed:

6.3.1.1 AutoML Approach

We implement our approach by building on the state-of-the-art AutoML framework AutoGluon [259] (`https://auto.gluon.ai/stable/index.html`). AutoGluon, developed by Amazon Web Services AI (AWS AI), is a tool for automating ML and DL model selection and training. In our pipeline, we use Multimodal AutoGluon (M-AutoGluon) as a neural AutoML solution to select the deep feature-extraction model, and Tabular AutoGluon (T-AutoGluon) to search over and select among traditional machine-learning classifiers. Specifically, after an initial deep-learning–based feature-extraction phase, the classifier chosen by T-AutoGluon is used to identify the type of roof maintenance issue. If an issue is detected, a second T-AutoGluon-selected classifier is then applied, for each type of issue, to estimate its severity.

Unlike many AutoML systems, AutoGluon does not rely on an explicit pipeline search or extensive hyperparameter-optimization loop. Instead, it trains a predefined set of model families and combines them using a multi-layer stacking and ensembling strategy.

In particular, T-AutoGluon adopts a three-layer ensemble architecture. In the first layer, multiple model families are trained on the same dataset. In the second layer, these base models are stacked and their predictions are used as inputs to higher-level models. Finally, the third layer aggregates the second-layer outputs through an ensemble selection method [260], a strategy also employed in Auto-Sklearn [261]. As with other AutoML frameworks, AutoGluon primarily aims to optimize model selection and predictive performance. In addition, it is designed to manage computational resources efficiently: it can early-stop iterative learners and selectively skip training some candidate models when they are unlikely to provide meaningful gains. When more training time

is available, AutoGluon can fit additional models and expand the ensemble, which often improves generalization.

A similar strategy is adopted by M-AutoGluon, which explores different backbones from the PyTorch Image Models (timm) library, along with associated hyperparameters and alternative ways of leveraging the backbone (e.g., full fine-tuning, partial layer freezing, or head replacement). However, as will be shown in the evaluation section, this approach has limited energy efficiency, as is common for methods that rely on training deep networks. Our idea is therefore to combine the two paradigms: M-AutoGluon is used to obtain a deep model that performs feature extraction, while T-AutoGluon is used to automatically select the most suitable traditional classifier on top of those features. Concretely, M-AutoGluon searches for the best CNN configuration (within PyTorch) until either an operator-defined time budget is reached or the search converges. By constraining the M-AutoGluon training time to the feature-extraction stage, we can obtain informative representations while delegating the final classification to T-AutoGluon.

In order to further accelerate the pipeline while maintaining strong predictive performance, the features extracted by the M-AutoGluon model can be compressed before being provided as input to T-AutoGluon.

6.3.1.2 Fuzzy-Transform Compression

In our methodology reported in Figure 6.9, we used the FT to compress the output from the drone and from the deep learning approach. In the following, we formally define the fuzzy-transform and will explain its application in our context.

Let $Y = [x_1, x_n]$ be a closed interval while $x_1 < x_2 < \ldots < x_n$ are points of S (nodes), with $n > 3$. It is possible to define a fuzzy partition of S as the sequence $\{P_j\}_{i=1,\ldots,n}$ of normal continuous fuzzy sets $P_i : S \to [0, 1]$, which satisfy the Ruspini condition $\sum_{i=1}^{n} P_i(x) = 1, \quad \forall x \epsilon S$. The fuzzy sets are generally known as basic functions. Such fuzzy partitions are uniform if the nodes are equidistant.

Common basic functions are the sinusoidal-shaped:

$$P_i(x) = \begin{cases} \frac{1}{2}\left(\cos(\pi\frac{x-x_i}{x_{i+1}-x_i}) + 1\right), & x\epsilon[x_i, x_{i+1}] \\ \frac{1}{2}\left(\cos(\pi\frac{x-x_i}{x_i-x_{i-1}}) + 1\right), & x\epsilon[x_{i-1}, x_i] \\ 0, & otherwise. \end{cases} \tag{6.18}$$

where x is the Input variable; x_i is the center of the i-th fuzzy set; x_{i-1} is the left neighbor node of x_i, x_{i+1} is the right neighbor node of x_i.

Another basic function is the gaussian-shaped:

$$P_i(x) = \exp\left(-\frac{(x - x_i)^2}{2\sigma^2}\right) \tag{6.19}$$

with x_i the Center of the i-th fuzzy set; σ the Standard deviation controlling the width of the bell curve.

While the pi-shaped is the following:

$$P_i(x) = \begin{cases} 0, & x \leq x_{i-1} \text{ or } x \geq x_{i+1} \\ 2\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right)^2, & x_{i-1} < x \leq \frac{x_{i-1}+x_i}{2} \\ 1 - 2\left(\frac{x - x_i}{x_i - x_{i-1}}\right)^2, & \frac{x_{i-1}+x_i}{2} < x < x_i \\ 1, & x = x_i \\ 1 - 2\left(\frac{x - x_i}{x_{i+1} - x_i}\right)^2, & x_i < x \leq \frac{x_i+x_{i+1}}{2} \\ 2\left(\frac{x - x_{i+1}}{x_{i+1} - x_i}\right)^2, & \frac{x_i+x_{i+1}}{2} < x < x_{i+1} \end{cases} \tag{6.20}$$

with x_{i-1} the start of rising slope; x_i the center of the i-th fuzzy set; x_{i+1} the end of falling slope.

Let us consider the $N \times M$ data matrix $\mathbf{X}$ and the fuzzy partitions $\{P_1, \ldots, P_n\}$ and $\{G_1, \ldots, G_m\}$ of the intervals $[1, N]$ and $[1, M]$ respectively, with $n < N$ and $m < M$. Let $\mathbf{P}$ and $\mathbf{G}$ be the matrices with entries $P_k(i)$ and $G_l(j)$, respectively.

The discrete F-transform matrix of $\mathbf{X}$ with respect to the fuzzy partitions $\{P_1, \ldots, P_n\}$ and $\{G_1, \ldots, G_m\}$ is given by $\mathbf{F} = \mathbf{H} \circ \mathbf{Q}$, where $\circ$ represents the Hadamard product, $\mathbf{H} = \mathbf{P}^T\mathbf{X}\mathbf{G}$ and $\mathbf{Q}$ is the matrix whose entries are the inverse of the entries of the matrix $\overline{\mathbf{Q}} = \mathbf{P}^T\mathbf{S}_{NM}\mathbf{G}$, with $\mathbf{S}_{NM}$ being the $N \times M$ matrix with all unit entries [262].

Each of these functions has been implemented as an algorithm that, starting from two-dimensional matrices, produces a compressed feature matrix $F \in \mathbb{R}^{n \times m}$. More details about the FT method and its pseudo-code are reported in [263].

In [264], it has been shown that, under suitable conditions, the order of the learning error obtained when using FT-compressed data is the same as that achieved with the original data. These conditions concern the bandwidth of the matrix $\mathbf{P}$ and the activation function adopted in the neural-network–based learning system. Moreover, as discussed in [262], the argument of the activation functions in the first hidden layer of the original architecture can be interpreted either as a combination of transformed input data or as a combination of the original data through transformed weights. In the latter case, the convergence of the learning system is tied to conditions on the initialization of the weights, requiring a unit ℓ_2-norm.

In Section 6.3.2, we present a case study that evaluates the complete methodology shown in Figure 6.9 on a real-world scenario based on the archaeological site of Pompeii.

6.3.2 Evaluation

As for FAUNO, the evaluation of HydraML was conducted through a case study on the archaeological site of Pompeii. In the following subsections, we describe the adopted datasets, the evaluation protocol, and the obtained results.

6.3.2.1 Dataset

The dataset was built starting from orthophotos of the archaeological site of Pompeii. We used nine orthophotos acquired between June 2023 and September 2024.

From these orthophotos, two roof datasets were extracted, corresponding to two roof typologies: (i) pitched tile roofs and (ii) flat roofs with conduits. In an archaeological site, structures are typically well known and evolve very slowly over time; therefore, it is feasible to deploy an autonomous drone to periodically collect images of these roofs.

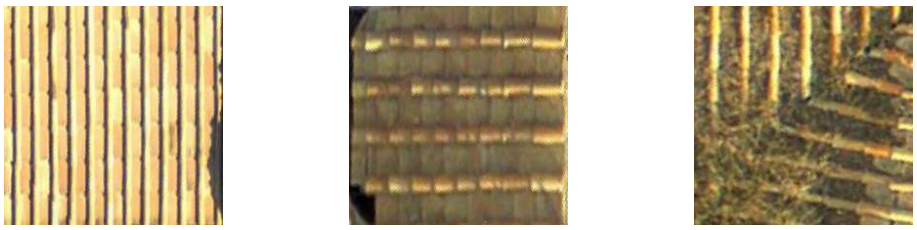
With the support of domain experts, roof classes were defined according to different types of issues. As shown in Figure 6.10, tile roofs can be categorized as normal (a), broken tiles (b), or weedy vegetation (c). Similarly, flat roofs (Figure 6.11) can be categorized as normal (a), broken conduits (b), stained (c), or weedy vegetation (d). Accordingly, we built two datasets: *PompeiiTile* and *PompeiiFlat*.

Overall, *PompeiiFlat* contains 922 images (347 normal, 111 broken, 283 stained, and 181 weedy), while *PompeiiTile* contains 1327 images (809 normal, 261 broken, and 257 weedy).

Since we do not have many images available, data augmentation was also performed. In particular, the data augmentation carried out consists of left right flip with probability 0.6, a gaussian blur with sigma between 0.2 and 1.2, rotation between -45,45 degree, change of photo exposure 0.4, 1.4, addition of Gaussian, Salt and Pepper and Poissonian noise. Data augmentation was performed to have a balanced training and validation sets during the cross-validation, in which each train class has 1200 samples and each val 300 samples. A detail worth noting is that such data augmentation was performed on the separate train and val for each fold after the dataset was split. For this reason, there are no augmented samples of the train in the val and vice versa.

Since our approach aims not only to classify roof issues but also to assess

their severity, we created additional datasets for severity prediction. Specifically, for each issue class, *broken*, *stain*, and *weedy* in *PompeiiFlat*, and *broken* and *weedy* in *PompeiiTile*, we built a dedicated dataset with three severity levels: *small*, *medium*, and *high*. For each issue-specific dataset, data augmentation was applied to obtain a balanced split with 500 samples per severity level for training and 100 samples per severity level for validation.



(a) Normal (b) Broken (c) Weedy

Figure 6.10: Tiles Roofs

(a) Normal (b) Broken (c) Stain (d) Weedy

Figure 6.11: Flat Roofs

6.3.2.2 Evaluation Protocol

We compare our approach against existing DL and AutoML baselines on both the *PompeiiTile* and *PompeiiFlat* datasets, as well as on the severity datasets associated with each roof type and issue category. The proposed method is implemented in Python and released as an open-source GitHub repository (`https://github.com/grosar/hydraml`).

The hardware used for our experimentation is the following:

- CPU: 13th Gen Intel(R) Core(TM) i9-13900K, 3000 MHz, 24 core, 32 logic processor;
- RAM: 64 GB;

- GPU: RTX A6000 with 48 GB of dedicated RAM.

The experimental phase is carried out using a 10-fold cross-validation protocol on both datasets. Following the analysis of related work in Section 6.1.2, we compare our method, in terms of both predictive performance and resource consumption, against recent CNN-based approaches for damage classification [239,242], as well as against DenseNet201, which is the backbone adopted in [242]. In addition, since our goal is to assess the effectiveness of the proposed hybrid strategy with respect to purely DL-based solutions, we also compare against M-AutoGluon, evaluating whether our approach can improve both performance and energy consumption.

We set a training time for each approach in the literature that is sufficient to reach convergence: 150 s (15 epochs), 480 s (30 epochs plus PCA and SVM), and 460 s (30 epochs), corresponding respectively to Wang et al. [239], Kumari et al. [242], and DenseNet201 on *PompeiiTile*. For *PompeiiFlat*, the same methods require 180 s (15 epochs), 690 s (30 epochs plus PCA and SVM), and 675 s (30 epochs).

For both datasets, we compare M-AutoGluon using a training time budget of 300 s for *PompeiiTile* and 180 s for *PompeiiFlat*. We also evaluate HydraML in the configurations based on Cosine, Gaussian, and Pi-shaped functions, with training times of 320 s for *PompeiiTile* and 200 s for *PompeiiFlat*. The additional 20 s accounts for the compression step and the training time of T-AutoGluon, which identifies the most suitable traditional classifier from the embeddings produced by M-AutoGluon.

Since the embeddings extracted by M-AutoGluon have 768 features, they are reshaped into $24 \cdot 32$ matrices and compressed into $12 \cdot 16$ matrices using the fuzzy transform. Reducing the number of features to less than half generally leads to an excessive drop in performance [262, 264].

DenseNet201 and VGG16 were trained using the Adam optimizer with $learning\ rate = 10^{-4}$, $\beta_1 = 0.900$, and $\beta_2 = 0.999$, while Kumari et al. [242] used PCA with $n_{\text{components}} = 100$ and an SVM classifier with a radial basis function (RBF) kernel.

The metrics used in our experimentation are the Accuracy (Acc), Precision (P), Recall (R), F1-Score (F1):

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{6.21}$$

Other metrics considered are $Precision_{macro}$, $Recall_{macro}$ and $F1_{macro}$:

$$\text{Precision}_{macro} = \frac{1}{K} \sum_{i=1}^{K} \frac{TP_i}{TP_i + FP_i} \tag{6.22}$$

$$\text{Recall}_{macro} = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FN_i} \tag{6.23}$$

$$\text{F1}_{macro} = \frac{1}{K}\sum_{i=1}^{K}\frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{6.24}$$

Given the number of classes K, TP_i is true positive, TN_i true negative, FP_i false positive, FN_i False negative and with $i \in \{1, \ldots, K\}$. The notation TP_i, TN_i, FP_i, FN_i means that such metrics are calculated when the problem is considered as binary with the samples of class i equals to the code 1 while the others will have the code 0.

In this paper, we also calculate the energy consumed by the approaches since our goal is to obtain good performance with also advantages in terms of sustainability: our goal is to show how the usage of T-Autogluon saves GPU training time. The average CPU and GPU power is calculated by sampling, during the training phase, the instantaneous CPU and GPU power every 100 ms. Let n be the number of acquired samples. The quantities CPU_{ipower} and GPU_{ipower} denote the cumulative instantaneous power, i.e., the summation of the n sampled values. The average power is therefore computed as:

$$CPU_{ipower} = \sum_{k=1}^{n} CPU_{ipower}^{(k)} \tag{6.25}$$

$$GPU_{ipower} = \sum_{k=1}^{n} GPU_{ipower}^{(k)} \tag{6.26}$$

$$CPU_{avgpower} = \frac{CPU_{ipower}}{n} \tag{6.27}$$

$$GPU_{avgpower} = \frac{GPU_{ipower}}{n} \tag{6.28}$$

Both the energy spent in a given time by GPU and CPU are calculated as

$$Energy = Power \cdot Time \tag{6.29}$$

6.3.2.3 Results

In the following, we report the results of our experiments. Table 6.4 shows that the best performance on *PompeiiTile* is achieved by the HydraML (Cos) configuration, with an accuracy of 0.9404, a precision of 0.9433, a recall of 0.9404, and an F1-score of 0.9406. This configuration also exhibits the lowest standard deviation across these metrics, which indicates higher stability with

respect to the other methods. For the *PompeiiFlat* dataset, Table 6.5 reports that HydraML (Cos) attains an accuracy of 0.9119, a precision of 0.9147, a recall of 0.9119, and an F1-score of 0.9117, again with the lowest standard deviation. These results indicate that the best performance is achieved by HydraML when using the cosine FT.

To clarify the advantages with respect to the M-AutoGluon baseline, it is important to motivate the adopted training time budgets for both M-AutoGluon and HydraML. These time budgets were selected because, as shown in Figures 6.12, 6.13, 6.14, and 6.15, they allow HydraML to match the converged performance of AutoGluon while requiring the shortest possible training time, resulting in substantial savings in overall training cost. With reference to Figure 6.12, we report the performance of M-AutoGluon and HydraML on *PompeiiTile* for different training time budgets, ranging from 120 to 480 s with a step of 60 s. The figure highlights that HydraML outperforms M-AutoGluon. In particular, at 300 s HydraML already achieves better results than M-AutoGluon trained for 160 s longer, when also accounting for the additional 20 s required to train T-AutoGluon. Specifically, HydraML reaches an accuracy of 0.9404 in Figure 6.12(a) and an F1-score of 0.9406 in Figure 6.12(b), whereas M-AutoGluon attains 0.9349 accuracy and 0.9350 F1-score. Although HydraML yields slightly higher scores at 420 s, we emphasize the 300 s setting to show that it surpasses a longer-trained M-AutoGluon with a smaller time budget. Overall, Figures 6.12(a) and (b) show that HydraML outperforms M-AutoGluon at all considered training-time intervals. Figure 6.13 reports the energy consumption, measured in joules and Wh, for the CPU (a) and GPU (b). The results indicate that reducing the training time budget from 480 s to 300 s enables HydraML to save approximately 40.92% of the GPU energy. In particular, M-AutoGluon consumes 109,993.3 J at 480 s, while HydraML consumes 64,975.6 J at 300 s.

A similar trend is observed for *PompeiiFlat* in Figure 6.14, where we report the performance of M-AutoGluon and HydraML for training time budgets ranging from 120 to 300 s with a step of 60 s. Also in this case, HydraML consistently outperforms M-AutoGluon. In particular, with 180 s of training HydraML surpasses M-AutoGluon trained for 100 s longer, achieving an accuracy of 0.9119 in Figure 6.14(a) and an F1-score of 0.9117 in Figure 6.14(b), compared to 0.9069 accuracy and 0.9068 F1-score for M-AutoGluon. Overall, Figures 6.14(a) and (b) confirm that HydraML outperforms M-AutoGluon at every considered training-time interval. Figure 6.15 reports the energy consumption of the CPU and GPU in Joules and Wh. The results show that HydraML saves approximately 12.28% of CPU energy, decreasing from 3746.7 J for M-AutoGluon at 300 s to 3286.3 J for HydraML at 180 s. Similarly, HydraML reduces GPU energy consumption by about 43.64%, from 66,350.5 J for M-AutoGluon at 300 s to 37,095.4 J for HydraML at 180 s.

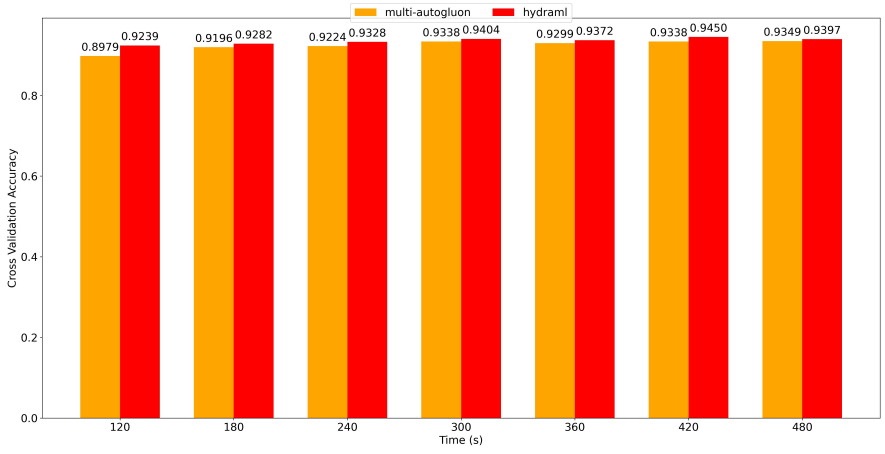
From Figures 6.16 and 6.17, we report the confusion matrices obtained from

Table 6.4: PompeiiTile Dataset Results

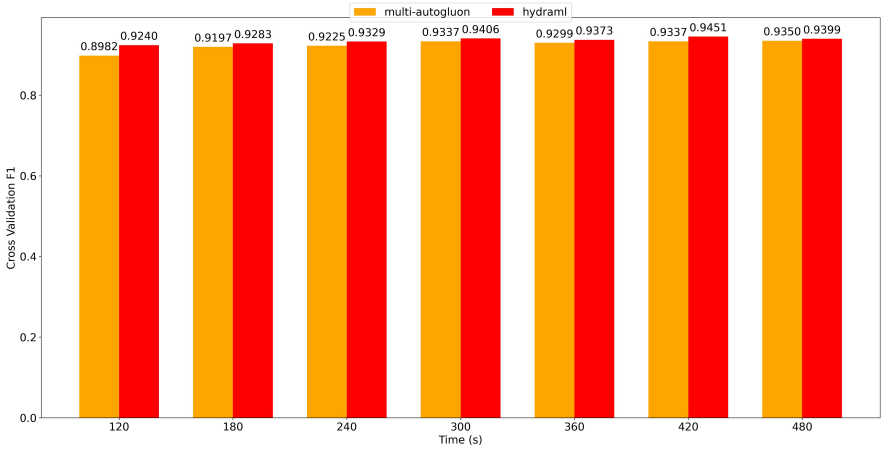
Model	Acc	P	R	F1	Time (s)	E CPU (J)	E GPU (J)	E CPU (Wh)	E GPU (Wh)
Wang et al. [239]	0.8769 (0.0288)	0.8821 (0.0253)	0.8769 (0.0288)	0.8774 (0.0280)	150	1152	31363	0.32	8.71
Kumari et al. [242]	0.8930 (0.0229)	0.8972 (0.0206)	0.8930 (0.0229)	0.8933 (0.0228)	480	4035	111228	1.12	30.90
DenseNet201	0.9157 (0.0198)	0.9190 (0.0184)	0.9157 (0.0198)	0.9159 (0.0198)	460	3333	104197	0.93	28.94
M-Autogluon	0.9338 (0.0243)	0.9358 (0.0227)	0.9338 (0.0243)	0.9337 (0.0244)	300	3541	64413	0.98	17.89
HydraML (Cos)	**0.9404 (0.0209)**	**0.9433 (0.0190)**	**0.9404 (0.0209)**	**0.9406 (0.0208)**	320	4056	64975	1.13	18.05
HydraML (Gaussian)	0.9373 (0.0250)	0.9397 (0.0236)	0.9373 (0.0250)	0.9376 (0.0249)	320	3954	64230	1.10	17.84
HydraML (Pishaped)	0.9266 (0.0232)	0.9304 (0.0203)	0.9266 (0.0232)	0.9269 (0.0229)	320	4038	62127	1.12	17.26

Table 6.5: PompeiiFlat Dataset Results

Model	Acc	P	R	F1	Time (s)	E CPU (J)	E GPU (J)	E CPU (Wh)	E GPU (Wh)
Wang et al. [239]	0.8774 (0.0406)	0.8816 (0.0383)	0.8774 (0.0406)	0.8771 (0.0408)	180	1407	37684	0.39	10.47
Kumari et al. [242]	0.8932 (0.0381)	0.8958 (0.0380)	0.8932 (0.0381)	0.8933 (0.0228)	690	5771	162943	1.60	45.26
DenseNet201	0.9080 (0.0307)	0.9109 (0.0296)	0.9080 (0.0307)	0.9074 (0.0318)	675	4770	153700	1.33	42.69
M-Autogluon	0.9069 (0.0346)	0.9098 (0.0332)	0.9069 (0.0346)	0.9068 (0.0346)	180	1608	36603	0.45	10.17
HydraML (Cos)	**0.9119 (0.0338)**	**0.9147 (0.0323)**	**0.9119 (0.0338)**	**0.9117 (0.0341)**	200	3286	37095	0.91	10.30
HydraML (Gaussian)	0.9018 (0.0418)	0.9061 (0.0405)	0.9018 (0.0418)	0.9014 (0.0421)	200	3096	37124	0.86	10.31
HydraML (Pishaped)	0.8968 (0.0388)	0.9006 (0.0380)	0.8968 (0.0388)	0.8966 (0.0392)	200	3399	37071	0.94	10.30

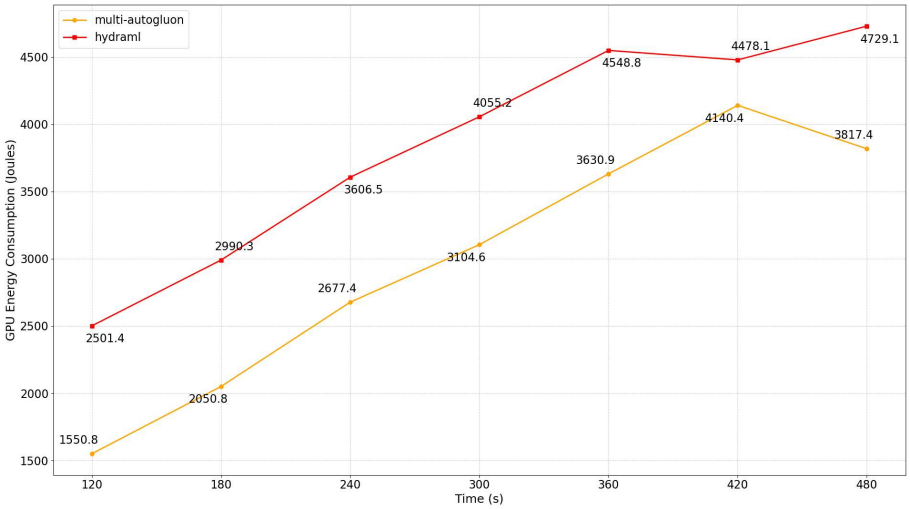


(a) Cross-Validation Accuracy

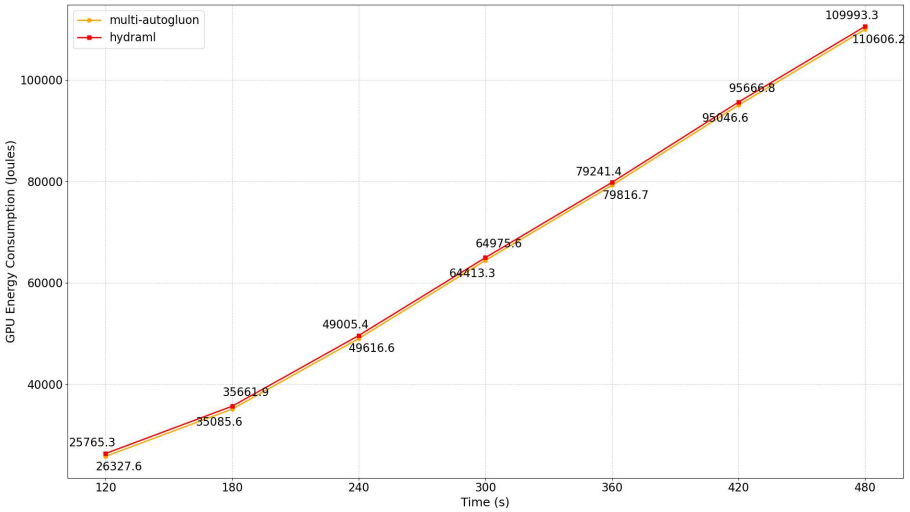


(b) Cross-Validation F1

Figure 6.12: PompeiiTile M-Autogluon and HydraML Cross-Validation Accuracy and F1 over Time of Training.

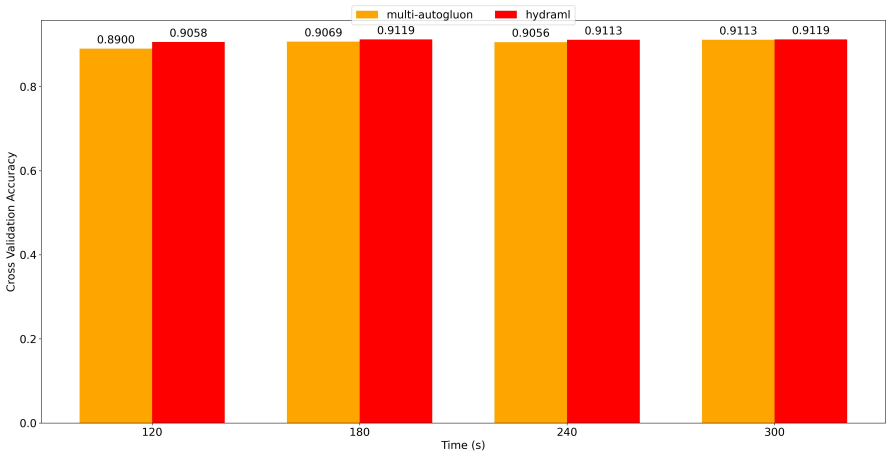


(a) Cross-Validation CPU Energy

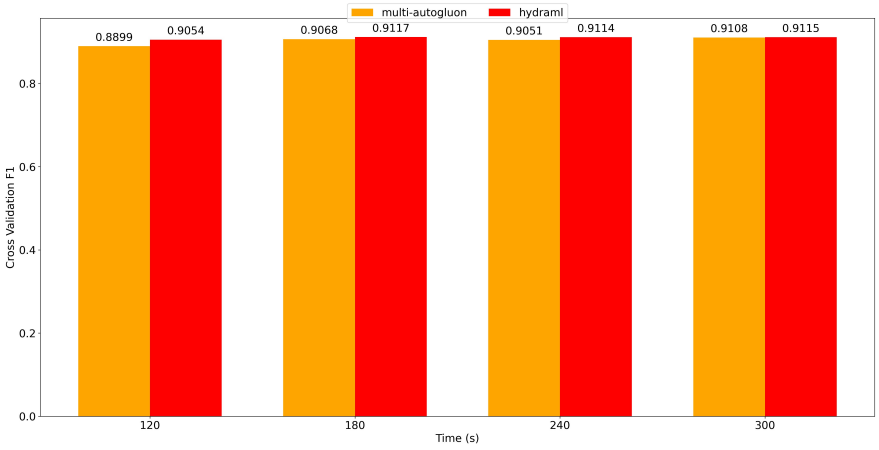


(b) Cross-Validation GPU Energy

Figure 6.13: PompeiiTile M-Autogluon and HydraML Cross-Validation CPU and GPU over Time of Training.

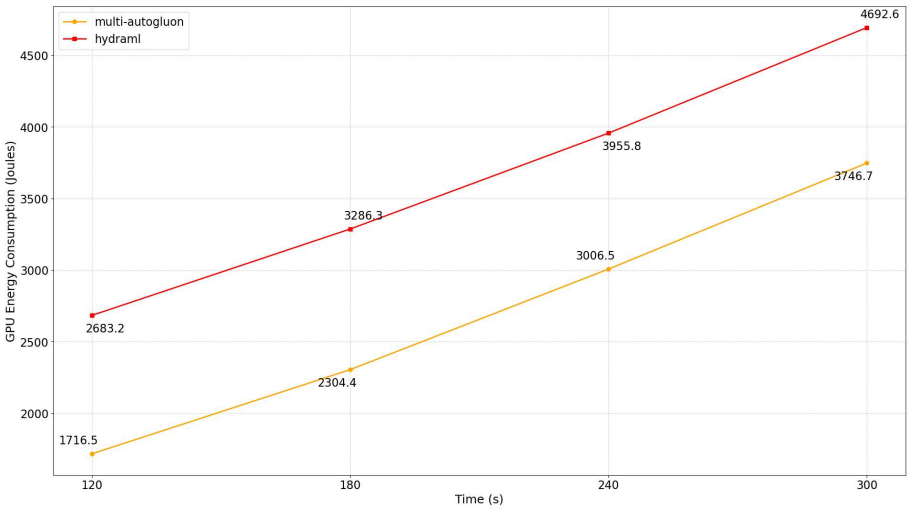


(a) Cross-Validation Accuracy

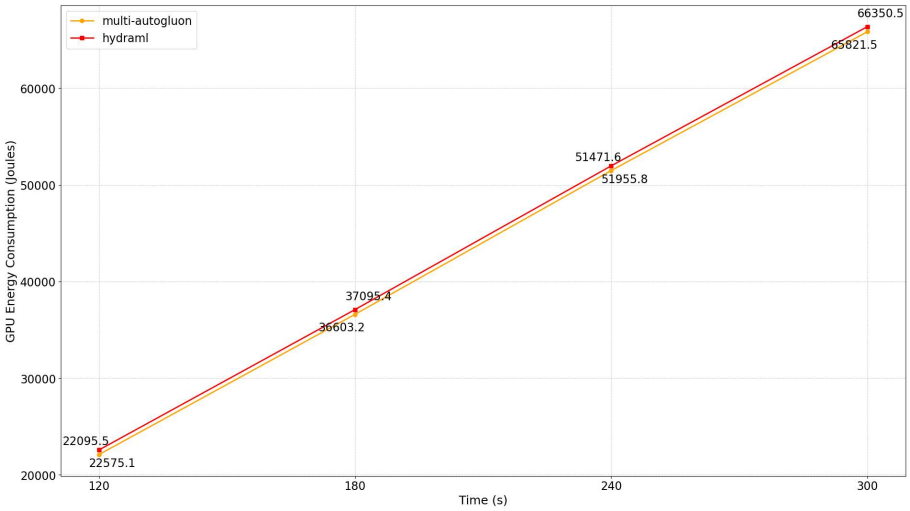


(b) Cross-Validation F1

Figure 6.14: PompeiiFlat M-Autogluon and HydraML Cross-Validation Accuracy and F1 over Time of Training.



(a) Cross-Validation CPU Energy



(b) Cross-Validation GPU Energy

Figure 6.15: PompeiiFlat M-Autogluon and HydraML Cross-Validation CPU and GPU over Time of Training.

the 10-fold cross-validation of *PompeiiTile* and *PompeiiFlat* using HydraML (Cos), trained for 320 s and 200 s, respectively. Each fold contains 300 samples per class. Since cross-validation is performed, we present the confusion matrices averaged across folds and expressed as percentages.

Figure 6.16, related to *PompeiiTile*, shows comparable correct classification rates for the *normal* and *weedy* classes, whereas a lower true-positive rate is observed for the *broken* class. In particular, 7.27% of *broken* samples are misclassified as *normal*. This behavior is expected, since some broken or missing tiles are visually subtle and may resemble mild discolorations or only slightly outlined tiles. A similar issue arises for *weedy* vegetation, which is misclassified as *normal* in 4.97% of cases. Notably, no *weedy* samples are confused with *broken* tiles, which suggests that the model relies on substantially different visual patterns for these two categories.

Figure 6.17, related to *PompeiiFlat*, shows the strongest performance for the *normal* and *broken* classes, while *stains* and, in particular, *weedy* vegetation are more challenging. Most *weedy* samples, amounting to 6.50%, are misclassified as *normal*, despite the presence of vegetation; the remaining errors are distributed between the *broken* and *stain* classes. In addition, 4.20% of *normal* samples are misclassified as *stain*.

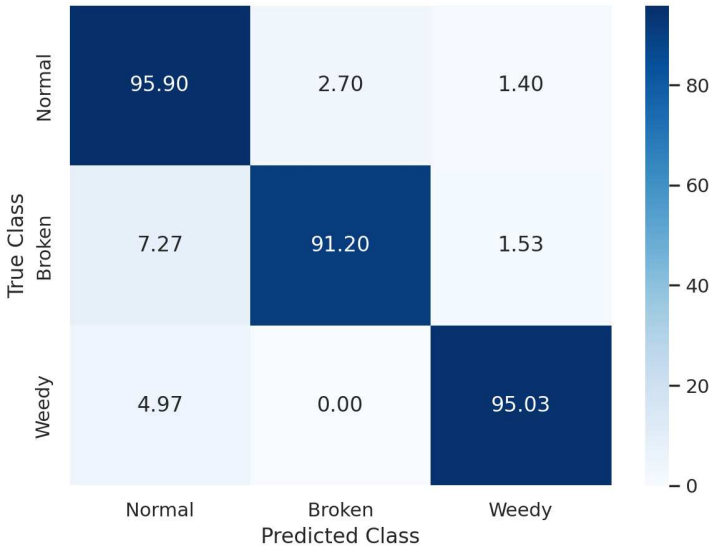


Figure 6.16: PompeiiTile Cross Validation Confusion Matrix (%)

A key strength of the proposed approach is that the hybrid combination of deep AutoML in M-AutoGluon and traditional classification in T-AutoGluon can be naturally extended to severity assessment. In particular, we reuse the feature embeddings extracted by M-AutoGluon to train additional classifiers

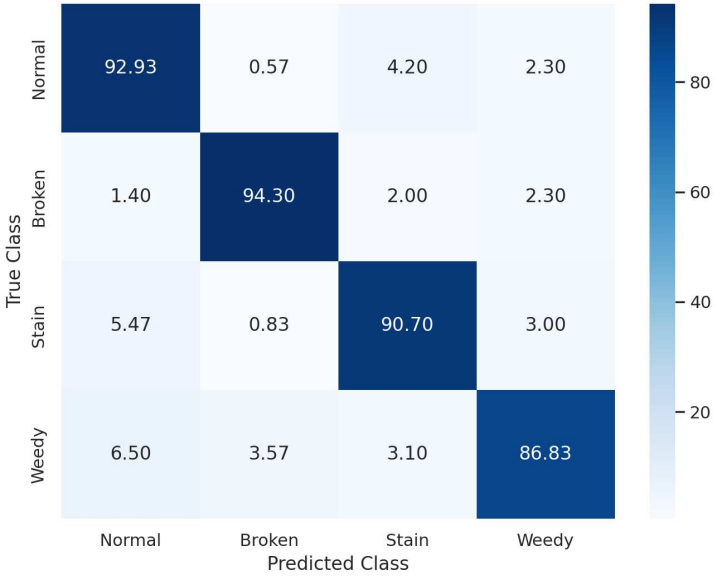


Figure 6.17: PompeiiFlat Cross Validation Confusion Matrix (%)

that estimate the severity of the detected issues. For each issue type, we define three severity levels, namely *small*, *medium*, and *high*.

Table 6.6 reports the results obtained with HydraML (Cos). For each severity sub-task, we train a dedicated T-AutoGluon model and report Accuracy, Precision, Recall, and F1-score, together with the CPU energy consumption in joules, since M-AutoGluon is not retrained. The results show that HydraML is highly effective in assessing the severity of broken tiles, achieving 0.97 across all metrics. Performance is around 0.89 for weedy vegetation on flat roofs, and around 0.85 for both broken tiles and weedy vegetation on pitched tile roofs. In contrast, the method is less effective for estimating the severity of damp patches on flat roofs, where the F1-score only slightly exceeds 0.75.

Additional insights can be obtained from the confusion matrices reported in Figures 6.18 and 6.19, which refer to severity classification on *PompeiiTile* and *PompeiiFlat*, respectively.

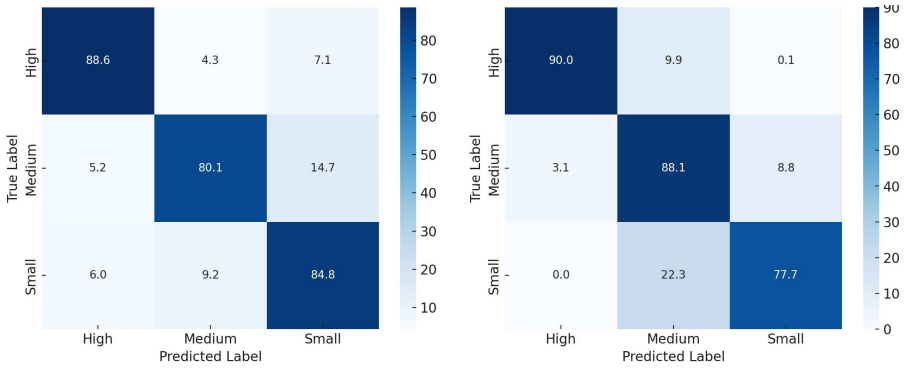
Figure 6.18 reports the confusion matrices for severity estimation on tile roofs, namely broken tiles (a) and weedy vegetation (b). For broken tiles, the most frequent errors occur between the *medium* and *small* classes, with 14.7% of *medium* samples misclassified as *small* and 9.2% of *small* samples misclassified as *medium*. This behavior can be attributed to the intrinsic difficulty of the task, since distinguishing between *small* and *medium* severity is challenging even at the annotation stage, and to the limited number of samples, which makes learning fine-grained differences more difficult. In addition, a non-negligible

Table 6.6: HydraML (Cos) Pompeii Dataset

Type of Roofs	Type of Issue	Acc	P	R	F1	E CPU (J)	E CPU (Wh)
Tile	**Broken**	0.8450 (0.0506)	0.8502 (0.0520)	0.8450 (0.0506)	0.8447 (0.0512)	1073	0.30
Tile	**Weedy**	0.8527 (0.0929)	0.8768 (0.0642)	0.8527 (0.0929)	0.8541 (0.0899)	1074	0.30
Flat	**Broken**	0.9744 (0.0312)	0.9748 (0.0309)	0.9744 (0.0312)	0.9745 (0.0311)	1080	0.30
Flat	**Stain**	0.7577 (0.0706)	0.7715 (0.0644)	0.7577 (0.0706)	0.7518 (0.0774)	1116	0.31
Flat	**Weedy**	0.8903 (0.0416)	0.8967 (0.0380)	0.8903 (0.0416)	0.8898 (0.0418)	1071	0.30

fraction of *high* severity samples is misclassified as *small*.

Similar considerations hold for weedy vegetation. In this case, 22.3% of *small* samples are misclassified as *medium*, and 8.8% of *medium* samples are misclassified as *small*. Moreover, the *medium* class is again critical, since approximately 10% of *high* severity samples are misclassified as *medium*.



(a) *Broken Tiles Severity Confusion Matrix (%)* (b) *Weedy Tiles Severity Confusion Matrix (%)*

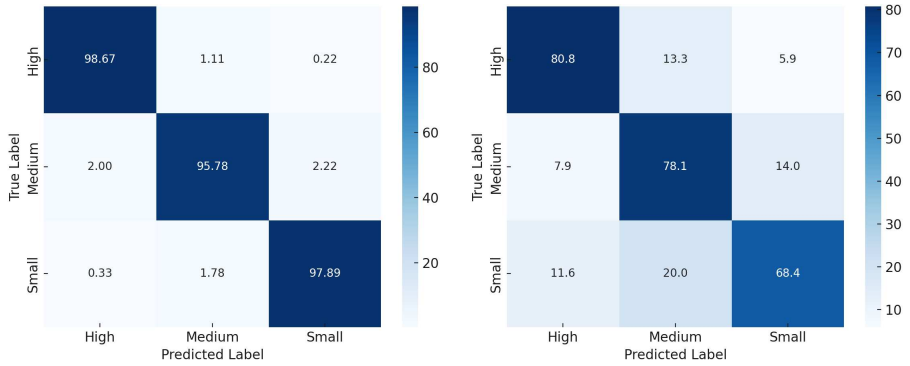
Figure 6.18: Tile Roofs Issues Severity.

Figure 6.19 reports the confusion matrices for severity estimation on flat roofs, namely broken conduits (a), stained conduits (b), and weedy conduits (c). For broken conduits, performance is consistently high across all classes. Most errors involve true *medium* samples, with 4% misclassified as *high* and 2% misclassified as *small*.

In contrast, the classification of severity for stained conduits is generally less accurate across all classes, with particularly low true-positive rates for

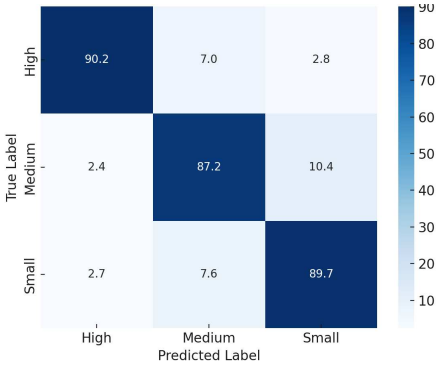
the *small* class. For weedy conduits, performance is instead satisfactory, with most errors concentrated between the *medium* and *small* classes. In particular, 10.4% of *medium* samples are misclassified as *small*, and 7.6% of *small* samples are misclassified as *medium*. Moreover, 7% of *high* severity samples are misclassified as *medium*.

Overall, the confusion matrices show that the majority of errors arise from confusion between *small* and *medium* severity, and more broadly from misclassifications involving the *medium* class. This trend suggests that the main difficulty lies in defining and recognizing *medium* severity, already at the ground-truth annotation stage.



(a) *Broken Conduits Severity Confusion Matrix (%)*

(b) *Stain Conduits Severity Confusion Matrix (%)*



(c) *Weedy Conduits Severity Confusion Matrix (%)*

Figure 6.19: Flat Roofs Issues Severity.

6.3.2.4 Discussion

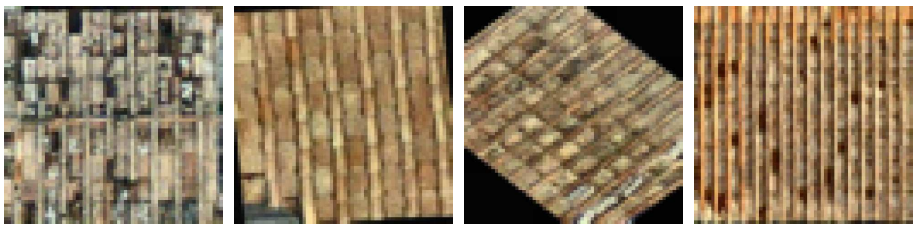
During the evaluation, our approach demonstrated clear advantages in terms of predictive performance, sustainability, and computational efficiency. The experimental results show that HydraML surpasses the approaches considered in the literature and their corresponding neural backbones, and it also outperforms M-AutoGluon. In addition, the proposed pipeline is designed to improve, or at least match, the performance of M-AutoGluon with shorter training time budgets. This is achieved by cascading T-AutoGluon on top of the embeddings extracted by M-AutoGluon, which enables stronger classification performance even when the deep component is trained for less time. Finally, the same embedding representations can be reused to train additional classifiers for severity assessment of the detected issues.

Despite these advantages, the proposed approach and the performed measurements also revealed limitations that are worth discussing to outline future research directions. Figures 6.18 and 6.19 highlight the overall behavior of the model across classes. However, it is also important to analyse which kinds of errors most frequently affect the predictions on both datasets. To this end, Figures 6.20 and 6.21 report representative examples of misclassifications.

Figure 6.20 presents typical errors on tiled roofs. In subfigure (a), a *normal* roof is misclassified as *broken*. This error is likely caused by dirt and mold, which generate visual patterns resembling missing or damaged tiles. Unfortunately, this situation is common and represents one of the main challenges to address, since roofs naturally deteriorate over time and local pixel patterns can be misleading. This issue could be mitigated by increasing the amount and diversity of training data, which is currently difficult due to the lack of public cultural-heritage datasets, apart from the one we release.

Subfigure (b) shows a roof with a clearly broken area at the bottom, which is nonetheless classified as *normal*. A plausible explanation is that the model focuses primarily on the central portion of the image, causing the damaged region to be underweighted and partially blended with the background. A similar effect appears in subfigure (c), where greenish patterns on the roof likely mislead the model into predicting *weedy*. Finally, in subfigure (d), the red patterns correspond to dry weedy vegetation. Due to the atypical color, this type of vegetation is harder to detect than green vegetation, and the model fails to recognize it correctly.

Figure 6.21 shows representative errors on flat roofs. In subfigure (a), a *normal* flat roof is misclassified as *stained*, due to the presence of a shadow that complicates the visual appearance and, consequently, the classification. Shadow-removal methods have been proposed in the literature, but they are typically developed for contexts that differ substantially from the ancient urban environment considered here. In addition, many of these techniques rely on

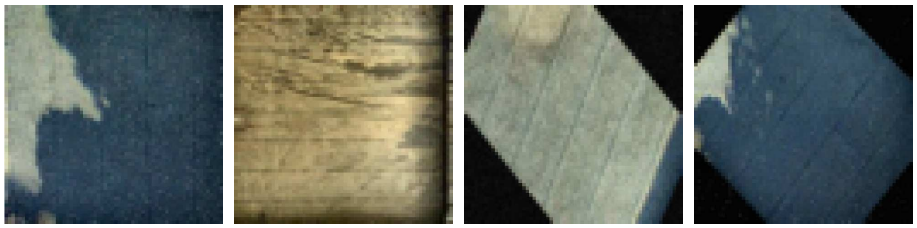


(a) Normal - Broken (b) Broken - Normal (c) Broken - Weedy (d) Weedy - Normal

Figure 6.20: Tiles Roofs Mistakes Examples (True - Predicted)

DL, which would increase both system complexity and computational cost. A similar issue is observed in subfigure (d), where a roof with *weedy* vegetation is misclassified as *stained*, again likely because of shadow effects.

In subfigure (b), a *broken* roof is misclassified as *stained*, plausibly because folded conduit patterns resemble moisture stain. In subfigure (c), a flat roof with a dust-related stain is misclassified as *stained* caused by dampness. This is a common source of error, partly because it can be challenging even for experts to determine whether a stain observed from above is due to moisture infiltration or simply surface dirt. This difficulty is analogous to the problem of damaged tiles, where it is often hard, even for trained annotators, to distinguish discolorations from truly missing tiles.



(a) Normal - Stain (b) Broken - Stain (c) Stain - Normal (d) Weedy - Stain

Figure 6.21: Flat Roofs Mistakes Examples (True - Predicted)

As illustrated by the examples in Figures 6.20 and 6.21, some errors can also be explained by the presence of a small number of ambiguous samples that exhibit visual cues typical of a different class. For instance, roofs affected by one type of issue may include small clusters of trees or vegetation, which resemble the *weedy* class and can mislead the model. Unfortunately, because these hybrid cases are relatively rare, the available data do not currently support the introduction of additional dedicated classes to explicitly model such situations.

6.4 Summary

The main goal of this chapter was to implement the perception and comprehension phase of the SA-CPSS shown in Figure 6.1, with particular focus on identifying maintenance issues at archaeological sites from aerial imagery. To this end, we proposed two alternative approaches: one aimed at detecting issues across the entire site, and another focused on classifying damaged roofs associated with structures that are already known within the archaeological area. The resulting knowledge about maintenance issues can then be used to estimate the vulnerability of site areas and to support the prioritization of interventions.

The first proposed approach, named FAUNO, is based on object detection with YOLO models and is intended for scenarios in which no prior information is available about the site or the location of its structures. Starting from orthophotos of the entire area, the model aims to detect all issues, which in our case are grouped into four categories: broken tiles, damaged conduits, weedy vegetation, and damaged structures. The detector therefore jointly estimates the location and the class of each issue across the site, producing bounding boxes together with their labels. Overall, FAUNO did not achieve satisfactory performance, especially for the *weedy vegetation* and *damaged structures* classes. Our analysis suggests that this is mainly due to the intrinsic difficulty of the detection task, which requires identifying damage patterns in an environment that is already naturally ruined and visually complex, and recognizing vegetation from an aerial viewpoint where shape, spatial arrangement, and texture can vary substantially. In particular, distinguishing trees from weeds is challenging when using only RGB aerial imagery. Hyperspectral data would likely provide more reliable vegetation identification [265, 266], enabling more accurate vegetation segmentation and facilitating the separation of weedy and non-weedy vegetation.

After acknowledging the difficulty of the detection task, we focused on the component that requires the highest level of accuracy, namely the classification of roof maintenance issues. Accordingly, assuming that the positions of the structures are known, we developed HydraML, an energy-aware approach that leverages automation to reduce training time and energy consumption while maintaining strong predictive performance. HydraML exploits AutoGluon in two stages. First, it uses the deep AutoML component of AutoGluon to select an effective model for feature extraction. The resulting embeddings are then compressed, using the Fuzzy-Transform, and used as input to a non-deep AutoML stage, again based on AutoGluon, to perform the final classification. Overall, HydraML achieves competitive results in terms of both accuracy and energy efficiency, outperforming other deep-learning baselines as well as the standalone deep AutoGluon configuration. HydraML will then be integrated

into the final prototype, which will be described and evaluated in Chapter 7.

Nevertheless, several directions for improvement remain. Future work includes devising methods to quickly and reliably estimate the most suitable modules, determining an optimal early-stopping point at which to halt deep training and switch to the downstream classifier, and increasing robustness to shadow-related effects.

Chapter 7

VERGIL: A SA Oriented Platform for Archaeological Site Protection

> "Can we get serious now? We've all heard about the computer simulations, and now we are watching actual sims, but I can't quite believe you still have not taken into account the human factor."
>
> — Chesley Sullenberger, *Sully*

The previous chapters have served to introduce, step by step, the various components necessary to implement the CPSS proposed in Chapter 4. However, to show their true usefulness with respect to our objective of supporting archaeological site operators, these modules must be integrated into a final system that can be used by the operators themselves. The CPSS must be made usable through the implementation of an application module (Figure 7.1) with an interface developed according to user centered criteria, in order to effectively support the operators' situation awareness.

As mentioned in the background chapter, these user-centered development principles focus on creating a system that is not primarily centered on the use of new technologies, but rather on helping users perform their tasks and on countering the so called demons of SA, particularly data overload and attentional tunneling.

Building on the CPSS methodological approaches and on the methods and techniques described so far, which are designed to be broadly applicable across CH contexts and beyond, we specialized them during the evaluation phase to test them at the archaeological site of Pompeii, developing the VERGIL

platform.

The design, implementation, and evaluation of VERGIL were carried out following a series of steps:

1. We conducted a Goal-Directed Task Analysis (GDTA) to identify operators' goals, decisions, and SA requirements.
2. The SA requirements were used to understand which data users need to make their decisions, what they need to understand, and how the identified situations may evolve in the future. This also made it possible to determine how to present this information. In such a way, the design of the proposed interface was created.
3. Starting from such interface design, the VERGIL implementation has been described.
4. The platform's ability to support user SA was evaluated using the SAGAT approach through user evaluation questionnaires administered after their involvement in various usage scenarios.

All these steps are described in detail in the following section:

Parts of this chapter have been previously published in:

- Rosario Gaeta, Giuseppe D'Aniello, and Vittorio Zampoli. *Situation-aware Adaptive Interfaces for Cultural Heritage Based on Oscillatory Attention Dynamics*. IEEE CH. 2025.
- Rosario Frontino, Rosario Gaeta, Rosalba Mosca. *Mining Knowledge from Data: The Case of Ontology Learning*. WIRN 2023. https://doi.org/10.1007/978-981-96-0994-9_30.
- Fabio Clarizia, Massimo De Santo, Rosario Gaeta, and Rocco Loffredo. *Enhancement Large Language Models Domain through Ontology-Based Retrieval-Augmented Generation.* International Journal on Semantic Web and Information Systems. 2025. Volume 21. https://doi.org/10.4018/IJSWIS.392507.

7.1 Background

First, we briefly introduce the theoretical background related to Goal-Directed Task Analysis, which is necessary for the design of the proposed platform in

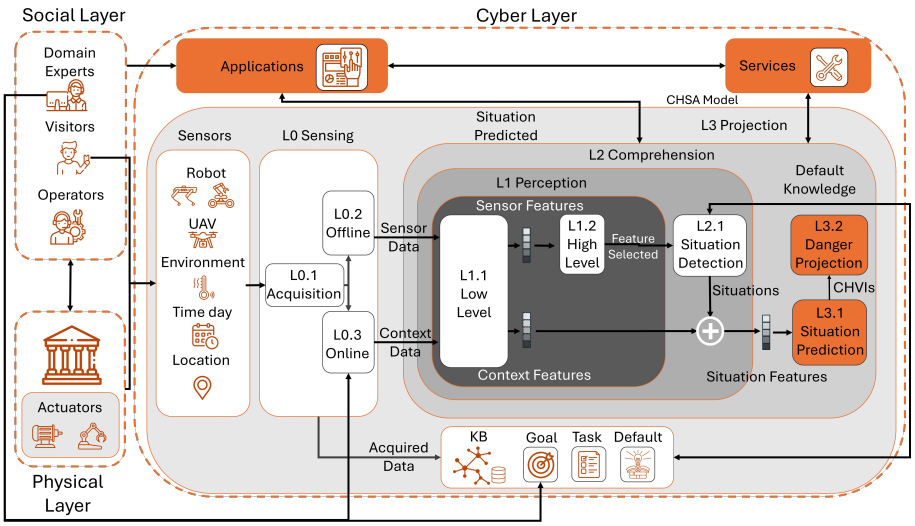


Figure 7.1: SA-CPSS with Projection and Application Details

the Pompeii Archaeological Site domain, and to Situation Awareness Measurement, which is required to understand and perform the evaluation of the platform.

7.1.1 Goal-Directed Task Analysis

GDTA is an approach used to identify both the user's major goals and the related situation awareness information requirements, and it is a form of cognitive analysis that is often applied in the initial stages of designing a situation awareness system [59].

GDTA is technology independent and it is used to identify the dynamic information that operators need in order to make decisions, without focusing on how they currently obtain it. The way information is acquired can vary widely across operators and systems, and it can change over time as technology evolves. By focusing on what information is needed rather than how it is obtained, designers can better understand how to support situation awareness, instead of simply reproducing the information displays of the current system. GDTA also helps clarify what operators should know to achieve each goal, even when that information is not yet available with current technology. If situation awareness requirements are defined only on the basis of today's tools, important information that operators would ideally like to have may remain overlooked.

To carry out a GDTA, experienced operators must be interviewed in order to identify goals, decisions, and situation awareness requirements. The interview results should then be combined with insights from written materials, maintenance plans, and other relevant sources to produce an initial GDTA. This initial version is validated by a larger group of experienced operators to ensure it captures all relevant goals, decisions, and SA requirements.

7.1.2 Situation Awareness Measurement

The evaluation of SA is a complex task, and within the scientific community the measurement of SA is still an open and debated issue. Some researchers argue that SA cannot be measured rigorously, while others, such as Endsley, maintain that its measurement is complex but possible, even with scientific rigour [58].

Indeed, the main difficulty lies in effectively measuring indices related specifically to SA rather than other factors involved in task performance, such as workload or attention, which only partially contribute to SA. It is therefore essential to clearly understand what is being measured and which measurement methods are used. A sound approach is to measure and quantify the improvement in SA levels resulting from the implementation of a new approach or the introduction of a new system feature within an existing system.

Indeed, the goal of measuring SA is to determine whether new approaches improve or degrade the user's SA. SA is often measured in relative terms by evaluating design concepts and approaches through comparative assessments. The relative improvement in SA is obtained by comparing a proposed approach with an existing one. There are typically no absolute thresholds for SA to be achieved; rather, higher levels of SA are generally preferable, as they indicate that the operator is more aware of ongoing situations and can respond more effectively. For this reason, relative measurement is the most natural and appropriate approach.

Numerous approaches for effectively measuring SA have been proposed. A performance index reflecting the level of SA is not sufficient on its own and must be considered together with the measurement context, the way in which any simulation is conducted, and the constraints of the testing situation, since these factors have a significant impact on the reliability and validity of the measurement.

In literature, the SA measurement methods are usually divided into the following four categories [58]:

1. **Direct system performance measures:** designing and executing evaluation scenarios in order to evaluate system performance, for example the time required to detect an anomaly. This involves introducing specific

anomalies to measure the operators' ability to detect them;

2. **Direct experimental technique:** surveys and probes are used to measure information seeking and represent the most common measurement approach, of which the Situation Awareness Global Assessment Technique (SAGAT) is one of the most widely used ones [23]. In this technique, the simulation is randomly suspended and the operator is asked a set of questions before the activity is resumed. Many other techniques have been derived from SAGAT. The main advantage of this approach compared to direct performance measures is that it allows the global assessment of SA, whereas performance based measures typically capture only specific aspects of SA through task outcomes;

3. **Verbal protocols:** video recordings of the observer during or after a simulation, exercise, or replay of situations are used. This approach is typically applied in the early stages of system evaluation and design;

4. **Subjective measures:** when objective metrics are difficult to obtain, it is possible to rely on self assessments, expert judgements, peer ratings, or supervisor ratings. A popular approach is the Situation Awareness Rating Technique (SART) [267], which estimates the user's SA by combining a set of sub scales into a single overall value. A limitation of this approach is that it tends to confound situation awareness with workload.

The choice of one measurement approach over another depends on the context in which it is applied, as well as on its relevance and usefulness. Endsley suggests which measures are most appropriate for different purposes [58]. As described above, verbal protocols are best suited for preliminary evaluations aimed at understanding SA requirements and identifying appropriate objective measures. Subjective measures should be used when no objective quantitative data are available or when objective measurement is impractical. Direct system performance measures can be useful for evaluating an individual's ability to satisfy specific SA requirements within a given scenario, or for assessing whether a system feature improves overall SA. Direct experimental techniques should be used to obtain objective measurements.

For our purposes, in order to evaluate improvements in SA provided by our VERGIL platform, we adopt the SAGAT approach.

The use of SAGAT gives us:

- an objective measurement of SA across all levels that is largely independent of the specific characteristics of the subjects involved;

- the reliability, sensitivity, and validity of SAGAT have been experimentally demonstrated in numerous studies [58];

- it is particularly suitable when the goal is to evaluate overall SA rather than a single specific aspect of a scenario;
- it supports the evaluation of relative comparisons between different versions of a system or its features, thereby assisting designers in the development of systems that better support SA.

SAGAT is a method that must be applied to real systems, applications, or prototypes and requires a considerable effort to design the surveys related to the SA requirements. However, it has the advantage of being highly reliable.

7.1.2.1 Situation Awareness Global Assessment Technique

SAGAT allows the assessment of global SA across all levels: perception, comprehension, and projection.

The technique requires that a real system be tested by operators and that its use be frozen at random times in order to ask users a set of randomly selected questions aimed at verifying their understanding of the situation. During this phase, the simulation is suspended until the users answer the questions within a given, randomly defined time interval. In this way, an objective measure of SA can be obtained, since the detailed information collected about the subjects' SA can be directly compared with the actual state of the system. Using this approach, SA is measured by directly assessing the operator's perceptions, rather than inferred from behavior, which could be influenced by many other factors.

The evaluation requires the definition of queries that are relevant with respect to the operators' SA requirements, as inferred from the GDTA. These SA requirements are those necessary to make the identified decisions and to achieve the related goals. The queries should probe all the SA requirements associated with these goals, be clear to the users, use appropriate domain specific terminology, be easy to answer, and be meaningful within the simulated scenario.

The procedure for conducting an evaluation using SAGAT should be defined according to the following recommendations provided by Endsley in [58]:

- The SAGAT procedures should be explained to the subjects before the testing phase;
- The normal tasks of the subjects should be performed while compile the SAGAT questionnaries should be considered a secondary task;
- While subjects are answering queries, they should only be able to see the questionnaire and no part of the system;

- The answering to the queries should have no penalty in case of errors, since the user should not be discouraged from answering queries. If a subject cannot answer to a query this provides important information in terms of SA;
- Subjects must not be placed in a position where they can interact with each other;
- For each scenario there is a fixed set of query but at each freeze of the simulation some random queries are selected;
- There may be some very important questions to ask during each freeze, but it is important that some queries are random;
- The moment when each freeze occurs should be random;
- Usually, no freezes should occur earlier than 3 minutes into a trial;
- Multiple freezes can happen in a trial. Three stops in a 15 minute trial can be a good rule of thumb;
- A freeze should last from 2 minutes to 5-6 minutes, regardless of the queries;
- The number of trials needed is related to the variability present in the scenario and the data samples of a trial. Usually, between 30 and 60 samplings per SA query across subjects and trials are good;
- A tolerance band is used to understand if each query is correct or incorrect.

7.2 Pompeii Archaeological Site Protection GDTA

In order to correctly design the VERGIL platform, it is necessary to identify the operators' goals within the archaeological site, the decisions they must make to achieve those goals, and the SA requirements needed to support those decisions.

The users' higher order goals required to perform the job must be identified. The decisions needed to achieve each goal are then listed under the corresponding goal. Next, SA requirements are defined by examining each decision in turn and specifying all the information the operator needs to make that decision. These information requirements should be stated without reference to any specific technology or method of acquisition, and they must encompass all the information necessary for decision-making across all three levels of SA.

In our case, the GDTA focused on the maintenance and protection of CH sites. It was developed through interviews with domain experts from the Pompeii archaeological site.

The system's macro-goal hierarchy is shown in Figure 7.2. The global aim of the project is to safeguard and protect archaeological sites while ensuring their revenue. Starting from this aim, two primary goals were elicited: *1.0 Ensure a quick and effective response to danger* and *2.0 Ensure the maximum economic income from visitors*. Both macro-goals have strategic economic value, with the first also being closely related to the preservation of the site's historical and cultural significance. However, these goals may be pursued concurrently, and progress toward one can hinder the other, thus requiring a careful balance in order to address both as effectively as possible. In this work, the focus is on maintenance strategy rather than revenue optimization; accordingly, the design of the system and the interface is grounded in the sub-goals associated with branch 1.0.

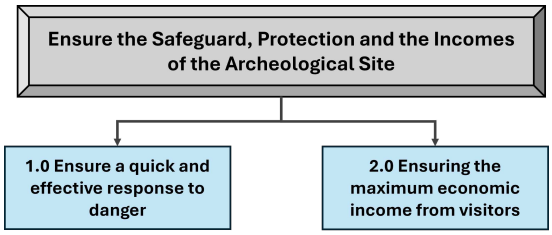


Figure 7.2: Macro Goal Hierarchy of Goal-Directed Task Analysis

The sub-goals of 1.0 are shown in Figure 7.3, where, for the first goal, the main decision is *How to update the maintenance plan?* This decision is particularly important because the maintenance plan guides all operations carried out on the site and therefore directly affects its overall condition. To answer this question, it is necessary to *1.1 Assess the maintenance issues of the structures*, *1.2 Assess the risks arising from environmental events*, and *1.3 Establish priorities for maintenance*. Each of these three sub-goals is essential for achieving the broader 1.0 goal. These goals are also partially concurrent: identifying new maintenance issues or monitoring environmental developments can complicate and slow down the process of setting priorities, since, at first glance, every issue may appear critical. It is therefore essential to pause and carefully analyse all available options in order to determine where intervention is most necessary.

Goal 1.1 is associated with sub-goals focused on identifying maintenance issues, specifically *1.1.1 Broken Tiles*, *1.1.2 Damaged Conduits*, and *1.1.3 Weedy Vegetation*. Goal 1.2 is associated with sub-goals related to contextual factors, in particular weather conditions and visitor presence. These correspond

to the goals *Assess the risks of weather events* and *Identify areas of risk for visitors*.

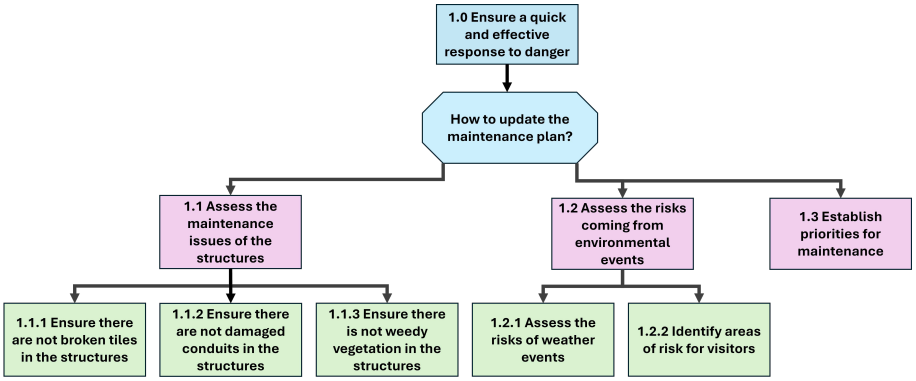


Figure 7.3: 1.0 Hierarchy Goal-Directed Task Analysis

From these sub-goals, we analyse the decisions that users must make and the situation awareness requirements needed to support those decisions.

The first part of the GDTA corresponds to branch 1.1, shown in Figure 7.4. In particular, the main decision associated with goal *1.1* is *Which structures need to be maintained?*, which can be addressed by answering the sub-decisions related to the specific maintenance issues, namely *In which structures do we need to address broken tiles, damaged conduits, or weedy vegetation?*.

Each of these sub-decisions is associated with a set of requirements aligned with the different levels of SA. In the perception phase, the user needs georeferenced aerial photos of the entire site, the locations of the roofs, and information on temperature, humidity, rainfall, and wind. In the comprehension phase, tiled and flat roofs must be identified, together with the presence of damage and weedy vegetation across the site. In the projection phase, the likely evolution of these conditions must be anticipated.

As can be seen from Figure 7.5, the goal *1.2* is connected to the decision *Which areas need to be secured?*, which highlights the competing nature of goals *1.0* and *2.0*. The need to secure areas by closing them to visitors, as required by this decision, conflicts with goal *2.0*, whose objective is to maximize economic revenue, since this goal cannot be fully achieved when areas must be closed.

The decision related to goal *1.2.1* is instead *Which areas need to be protected against weather conditions?*. For this decision, at the perception level it is essential to know historical data, current conditions, and forecasts for temperature, humidity, rainfall, and wind, together with information about the materials of the structures and their vulnerability to erosion and damage. At the comprehension level, it is necessary to combine the information on broken

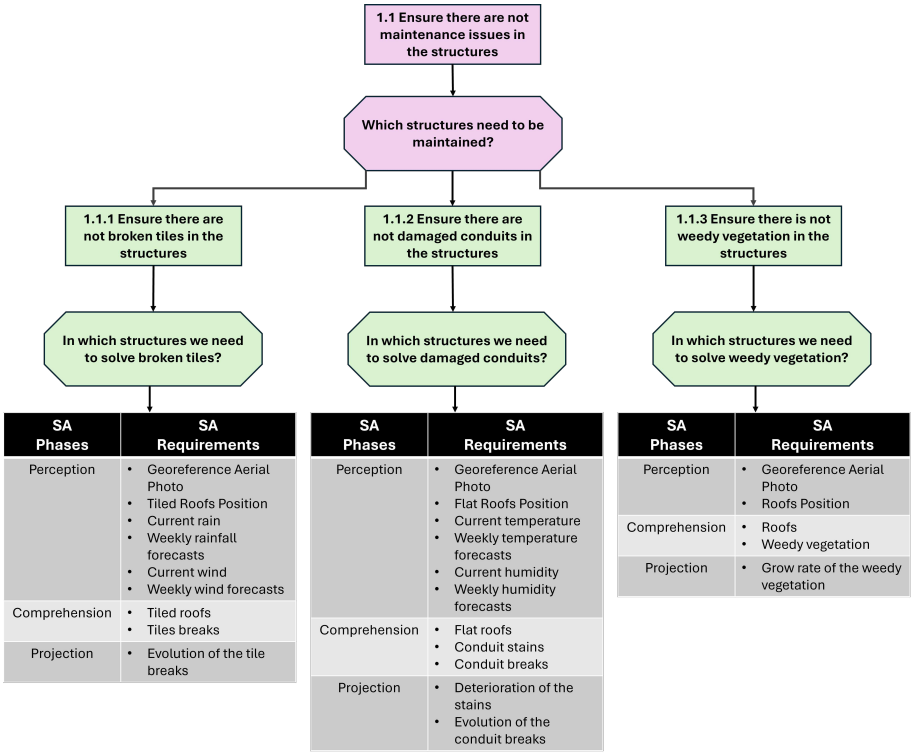
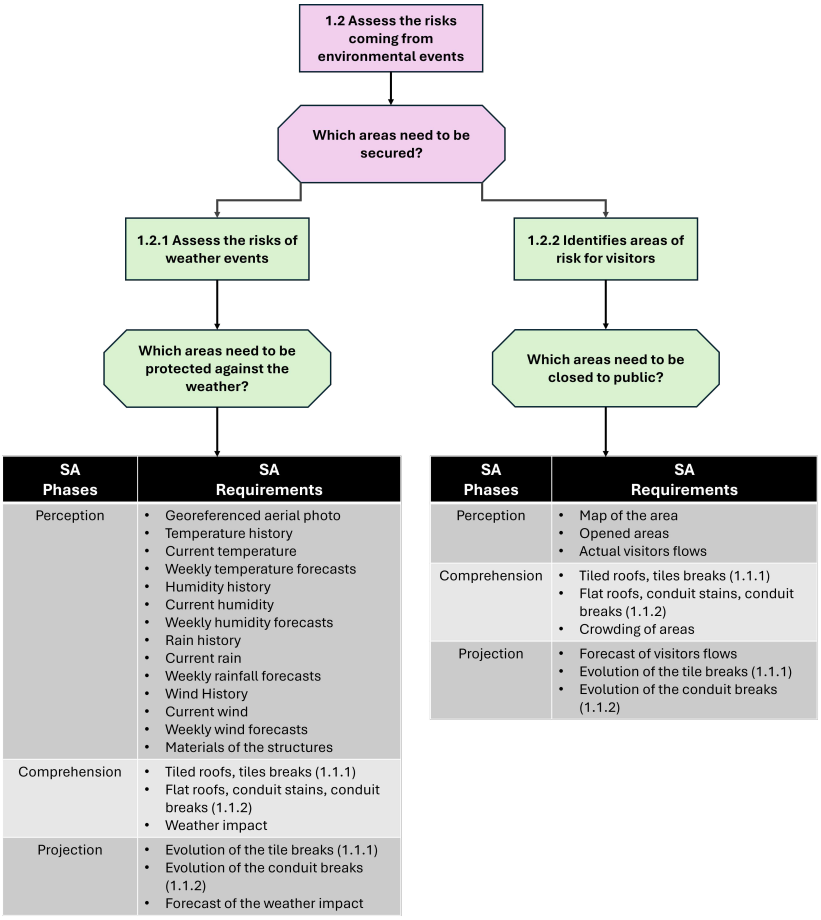


Figure 7.4: 1.1 Goals, Decisions and SA Requirements

tiles and damaged conduits identified under goal *1.1* with an understanding of weather impacts, in order to determine which areas require protection from environmental conditions. At the projection level, it is fundamental to anticipate how weather impacts may evolve over time, together with the consequences of broken tiles and damaged conduits.

Considering goal *1.2.2*, all SA requirements focus on identifying which areas should be closed to the public, either because they pose a risk to visitors or because visible damage should be avoided. At the perception level, information about open areas and current visitor flows is required in order to identify crowded zones which, when combined with the broken tiles and damaged conduits identified under goal *1.1*, allows the user to understand the situation and determine whether an area should be closed. It is also necessary to understand how visitor flows and damage conditions change over time.

Finally, the last goal to be analysed is *1.3* (Figure 7.6), which is connected to the decision *Which areas require the most immediate intervention?*. While the other two main sub-goals focus on understanding the situation in terms of events or damage, this goal is concerned with identifying which situations demand greater attention and is therefore crucial for determining how the maintenance



SA Phases	SA Requirements
Perception	• Georeferenced aerial photo • Temperature history • Current temperature • Weekly temperature forecasts • Humidity history • Current humidity • Weekly humidity forecasts • Rain history • Current rain • Weekly rainfall forecasts • Wind History • Current wind • Weekly wind forecasts • Materials of the structures
Comprehension	• Tiled roofs, tiles breaks (1.1.1) • Flat roofs, conduit stains, conduit breaks (1.1.2) • Weather impact
Projection	• Evolution of the tile breaks (1.1.1) • Evolution of the conduit breaks (1.1.2) • Forecast of the weather impact

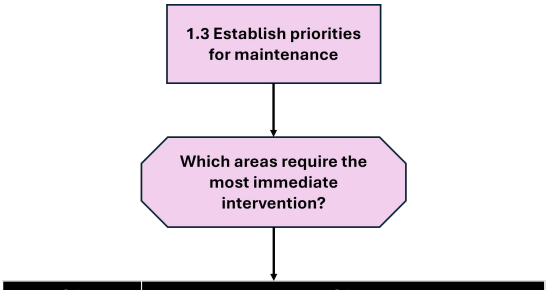
SA Phases	SA Requirements
Perception	• Map of the area • Opened areas • Actual visitors flows
Comprehension	• Tiled roofs, tiles breaks (1.1.1) • Flat roofs, conduit stains, conduit breaks (1.1.2) • Crowding of areas
Projection	• Forecast of visitors flows • Evolution of the tile breaks (1.1.1) • Evolution of the conduit breaks (1.1.2)

Figure 7.5: 1.2 Goals, Decisions and SA Requirements

plan should be updated.

At the perception level, the SA requirements include the current maintenance plan to be followed, the materials and construction techniques of the structures, their usage, prior interventions, decorative elements, water features, and vegetation species. To establish maintenance priorities across the different areas, it is necessary to integrate all the maintenance issues identified under goals *1.1.1*, *1.1.2*, and *1.1.3*, together with the contextual information related to goals *1.2.1* and *1.2.2*, as well as data on rain and wind accumulation and the time elapsed since the onset of adverse weather conditions. Taken together, these elements make it possible to assess the vulnerability of all structures, also through the use of specific indices, in order to identify which areas are in the poorest condition and therefore require the most urgent intervention. The user must also anticipate how these situations may evolve over time in order to

understand how vulnerability could change in the future.



SA Phases	SA Requirements
Perception	• Actual Maintenance plan • Construction and materials of the structures • Usage of the structures • Previous Intervention on the structures • Element of decoration of the structures • Water features in the structures • Species of vegetation in the structures
Comprehension	• Tiled roofs, tiles breaks (1.1.1) • Flat roofs, conduit stains, conduit breaks (1.1.2) • Roofs, weedy vegetation (1.1.3) • Weather Impact (1.2.1) • Crowding areas (1.2.2) • Rain cumulation • Wind cumulation • Time elapsed since the adverse rainfall and wind
Projection	• Evolution of the tile breaks (1.1.1) • Evolution of the conduit breaks (1.1.2) • Grow rate of the weedy vegetation (1.1.3) • Forecast of the weather impact (1.2.1) • Forecast of visitors flows (1.2.2) • Evolution of cumulated adverse rainfall and wind • Persistence of the adverse rainfall and wind

Figure 7.6: 1.3 Goals, Decisions and SA Requirements

7.3 VERGIL Adaptive Interface

The SA requirements defined in the previous section form the basis for the design of the VERGIL platform interface. Starting from the GDTA, it is possible to define an interface capable of presenting all the information required to support the different levels of SA in a user friendly manner. The platform implements the CPSS shown in Figure 4.1 and integrates all the discussed modules, including the application module.

7.3.1 Main Components and Implementation

The VERGIL platform has been implemented using the Django framework (`https://www.djangoproject.com/`), with the backend developed in the Python programming language and the frontend implemented in JavaScript. The backend provides the sensing, perception, comprehension, and part of the projection functionalities, implementing all the modules introduced in Chapters 5 and 6.

These features are accessible through an Adaptive Operators Interface designed to support users in performing their tasks. The interface developed for archaeological sites can be divided into two macro areas:

1. A map of the monitored area highlights the vulnerability of the different zones. The map displays the maintenance issues automatically detected using AI, in line with the GDTA Goal *1.1*, and shows the selected area according to Sub-goal *1.2.2*. It is designed to support the understanding of intervention priorities, as defined in Goal *1.3*;

2. Contextual information related to the area, as specified in Goal *1.2*, is displayed through various widgets.

The map within the interface consists of a high definition aerial orthophoto of the entire area under analysis. The map has been integrated using the Leaflet library for interactive maps (`https://leafletjs.com/`), adopting the Monte Mario 2 reference system and integrating a map of Regio I. The management of the map at the backend level and the georeferencing functionalities were handled using the GDAL (`https://pypi.org/project/GDAL/`) and Rasterio (`https://pypi.org/project/rasterio/`) libraries.

The map highlights the different sub-areas, referred to as insulae. There are 25 insulae in the region, of which 20 are relevant to this study. Each insula is associated in real-time with a CHVI, which is linked to a specific color that visually represents its level of vulnerability.

The CHVIs are computed by combining contextual data with maintenance issues detected in the orthophoto. When the orthophoto is loaded, previously mapped structures are cropped, processed, and provided as input to the HydraML technique, which assigns a class to each structure. Each type of damage is identified by a rectangular area, whose color indicates the type of issue, as described in the corresponding map legend. The counts of damage and weedy vegetation are then merged with the context related contributions and combined with the static component of the CHVI to produce the final CHVI visualization.

The color of each insula also has a specific meaning, as it is directly connected to the CHVI value, which ranges between 0 and 1. The level of vulnerability is therefore indicated by the color of the insula as follows:

- **Low - Green** ($0.00 \leq CHVI < 0.50$)
- **Medium - Yellow** ($0.50 \leq CHVI < 0.65$)
- **High - Orange** ($0.65 \leq CHVI < 0.85$)
- **Very High - Red** ($0.85 \leq CHVI \leq 1.00$)

Each insula is associated with a label reporting key information about the area, including the insula name, identified by an ID number, the CHVI value, and whether the insula is visitable or not. This information is essential for understanding visitor flows within each area. The overlay linked to this label also allows the user to view information related to the state of preservation and the materials of the insula, as reported in Tables 4.1 and 4.2. These data are retrieved from a knowledge base that will be described later, together with their representation in the user interface.

Figure 7.7: Overview Map of the SA Interface

Below the map, as can be seen in Figure 7.8, there is a section entirely dedicated to contextual data that can be perceived by human operators. In particular, a first strip of widgets is devoted to the five types of contextual information considered in this work: four related to weather conditions, namely temperature, humidity, wind, and rainfall, and one related to weekly visitor numbers. Each widget displays the current data recorded in real-time. Since the experiments refer to periods that have already occurred, specifically September 2023, these data are retrieved from the sources described in Chapter 5.

For each weather variable, a bar plot shows the current value together with the values from the previous seven hours and the forecasts for the next seven hours. This representation is intended to provide operators with both the recent

history and short term future evolution of these variables, which is particularly important given that weather forecasts are generally reliable in the short term but become less accurate over longer horizons. Following the same rationale, cumulative rainfall is also visualized, calculated using the formula given in Equation 5.1 described in Chapter 5.

The visitor bar plot instead shows the total number of visitors to the site for the current week and for the previous eight weeks.

All this information, especially when combined with photographs of the site and data on maintenance issues, supports site operators in understanding the current situation and in determining the most appropriate actions to take.

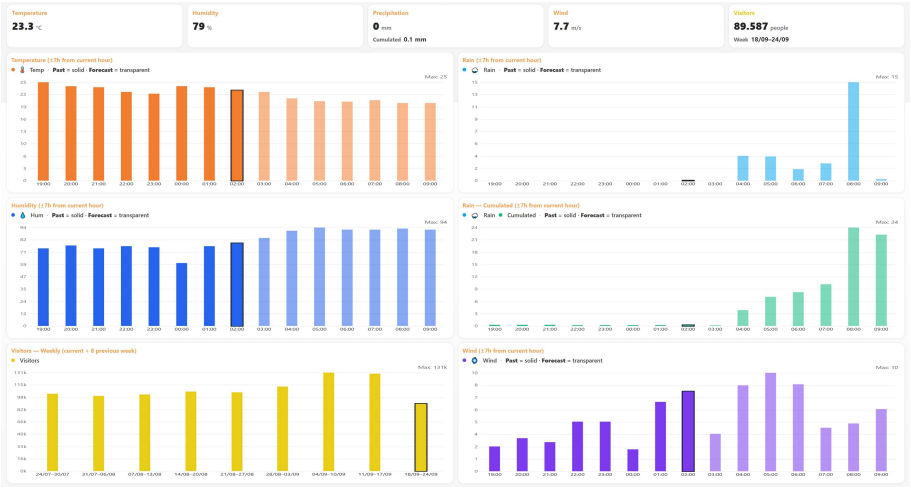


Figure 7.8: Contextual Data on the Interface

7.3.2 VERGIL Knowledge Base

For a user browsing the map of a specific area within the archaeological site, it is essential to understand the areas, structures, and characteristics of the elements being observed. Often, users already have some prior experience and are therefore familiar with parts of the site. However, in an area as extensive as a Regio I, and even more so across the entire site, it becomes difficult to recall all structures and their characteristics based solely on the map. When sufficient information is not readily available, even experienced users may rely on memory, which can lead to cognitive bias and potentially incorrect decisions.

For this reason, it is essential to provide a knowledge base that can be queried in order to access all necessary information when required. A knowledge base can be formally defined as a structured repository of concepts, entities, and

relationships, composed of data sources [268]. These sources often include ontologies, which are defined as formal and explicit specifications of a shared conceptualization and are used to support reasoning, querying, and decision-making [269]. Ontologies, however, are typically complex to build, which has led to the development of ontology learning techniques that enable their automatic or semi automatic creation from structured, semi structured, or unstructured data sources [270].

This knowledge base is made available through the dedicated module highlighted in Figure 7.9, which illustrates its role within the CPSS.

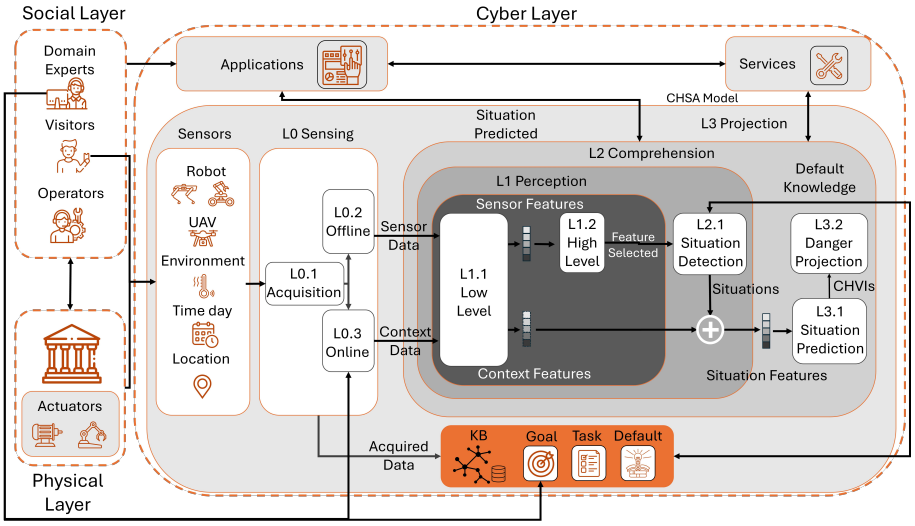


Figure 7.9: SA-CPSS with Knowledge Base Details

Despite existing research on ontology learning from SQL, which involves structured data, and NoSQL, which involves semi-structured data [271], to the best of our knowledge there are no publicly available tools that can be readily used in practice. For this reason, starting from algorithms that were partially defined in previous works, we refined and simplified their main limitations, making them usable by a wider audience through the development of two tools. The first tool focuses on SQL to OWL conversion and is distributed as open source software at `https://github.com/grosar/sql_to_owl_method`. The second tool performs JSON to OWL conversion and is provided as an online service at `https://json2owl.centroictbc.unisa.it/`.

In our case, the information required about the structures of Pompeii, their locations, and the characteristics of the structures within a given area is available in JSON format. For this reason, we describe the designed approach and the tool implemented to convert JSON data into an OWL ontology, named JSON2OWL. We also discuss the resulting ontology and its integration into the final interface,

highlighting the functionalities enabled by the use of this ontology.

7.3.2.1 JSON2OWL tool

The proposed approach enables the semi automatic conversion of populated JSON NoSQL databases into OWL ontologies. By selecting a JSON file together with a mapping file, domain experts can generate both the TBox, which represents the terminological component, and the ABox, which represents the assertional component, of an ontology. The TBox is generated using conversion rules for JSON data directly derived from [271] and is enriched with mechanisms that support active customization of class and property names, as well as the conversion between data properties and object properties. The ABox Generator module creates instances for all defined classes by parsing the database and assigning unique identifiers derived from the corresponding class names.

As described, two algorithms are responsible for the creation of the ontology schema and for its subsequent population, respectively.

The TBox approach is based on three fundamental rules:

- **Rule 1:** Each JSON object is converted into a class;
- **Rule 2:** If the values associated with a key are simple types, then we will have a dataProperty whose domain is the class corresponding to the object containing the key and the range is the specific simple type. The name of the dataProperty will be the concatenation of *has_a_* + KEY_NAME + *_of_* + DOMAIN_NAME;
- **Rule 3:** If an object is integrated into another, we will have an objectProperty whose domain is the external class and whose range is the internal class, which will then be created and named using the name found in the JSON. The name of the objectProperty will be the concatenation of *has_a_* + *INTERNAL_CLASS_NAME* + *_of_* + *EXTERNAL_CLASS_NAME*. This also applies to objects integrated into the main JSON object, named after the file name, for example Object.json.

The ABox generation approach exploits the data contained in each class, according to Rule 1, to instantiate individuals of that class, including data properties as defined by Rule 2 and object properties as defined by Rule 3. By default, the names of individuals are derived from the class name combined with a sequential identifier, which allows different instances of the same class to be distinguished. This typically occurs when multiple JSON objects share a similar or identical structure. The tool also supports instance renaming by associating a class with one or more data properties whose domain is either the

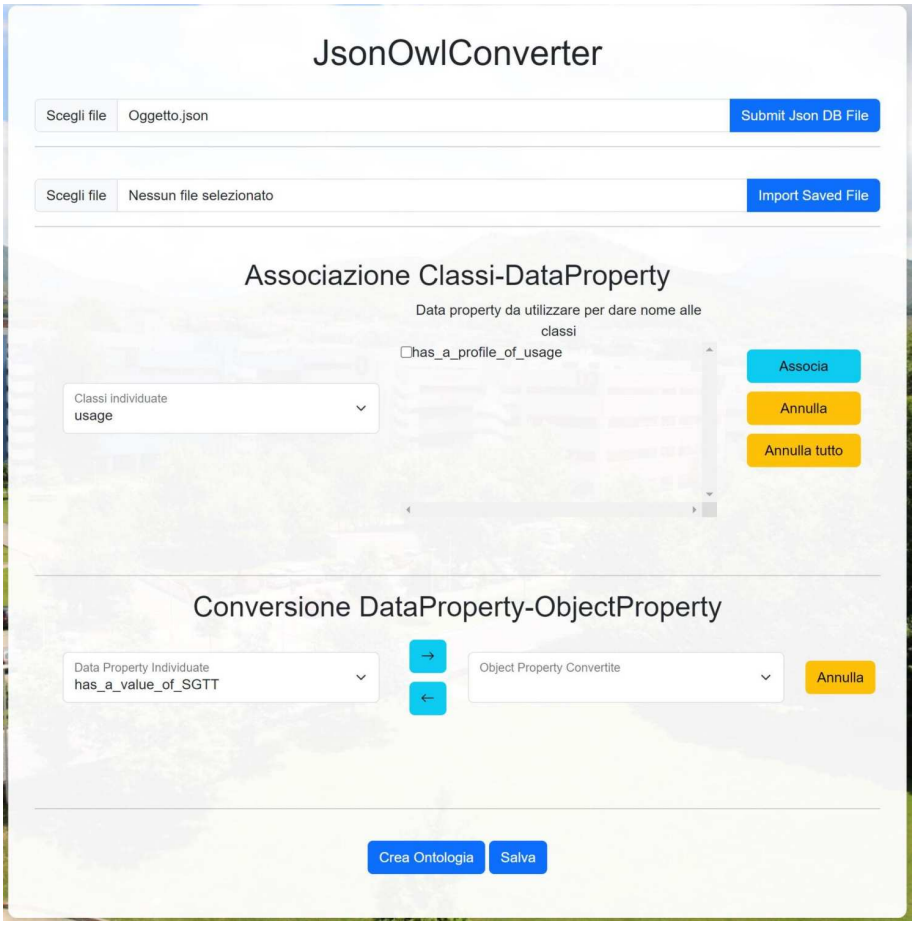


Figure 7.10: JSON2OWL Tool Interface: Page 1

class itself or an external class. In this case, the instance names are assigned using the values of the selected data properties.

Data types are also converted following a one-to-one translation as reported in Table 7.1. The proposed solution has been implemented as a Python web

JSON	**xsd/owl**
boolean	`xsd:boolean`
number (float)	`xsd:decimal`
number (int)	`xsd:integer`
string	`xsd:string`

Table 7.1: JSON to OWL Conversion Table

application using the flask framework https://flask.palletsprojects.com/en/stable/ and it is available at https://json2owl.centroictbc.unisa.it/.

The tool allows users to select a JSON file and submit it for conversion. One of its main purposes, as well as a key difference with respect to previously proposed approaches, including [271], is that it was explicitly designed as a semi automatic solution. Since certain pattern types, as also discussed in related work, cannot be translated accurately through a fully automatic process, the tool enables users to actively intervene during the conversion. In particular, users can modify names, create new concepts, and define additional object properties in order to refine the resulting ontology.

As shown in Figure 7.10, the Class DataProperty Association functionality allows instances of a class to be renamed by associating the class with one or more data properties whose domain is either the class itself or an internal class within the JSON structure. In this way, instance names are derived from the values of the selected data properties. To perform this operation, the user first selects the class whose instances are to be renamed from the *Identified Classes* drop down menu, and then selects the data properties to be used for renaming by means of the corresponding checkboxes. The association is applied by pressing the *Associate* button.

If the selected property is not consistently present across all JSON objects, the instance name defaults to a progressive identifier derived from the class name, and any other associations defined for that class are discarded. The DataProperty-ObjectProperty Conversion functionality allows a data property to be transformed into an object property, with the property to be converted selected from the drop down menu on the left. Users can also rename classes and properties when necessary.

7.3.2.2 Pompeii Archaeological Site Ontology

The JSON used as a data source was obtained by combining CSV data from other projects, such as https://github.com/p-lod/p-lod-csv, data from the official portal of the Pompeii archaeological site, and information collected from Pompeii domain experts who evaluated the state of conservation, materials, and other characteristics of the structures within the site.

The ontology obtained using JSON2OWL is shown in Figure 7.11 and consists of 10 classes, 8 object properties, and 30 data properties. The ontology is structured around the concept of *Regio*, which is composed of multiple insulae that divide the Regio into several sub-areas. Each insula is associated with criteria and sub criteria that describe its vulnerability and other characteristics related to the condition and structure of the area. In addition, each insula

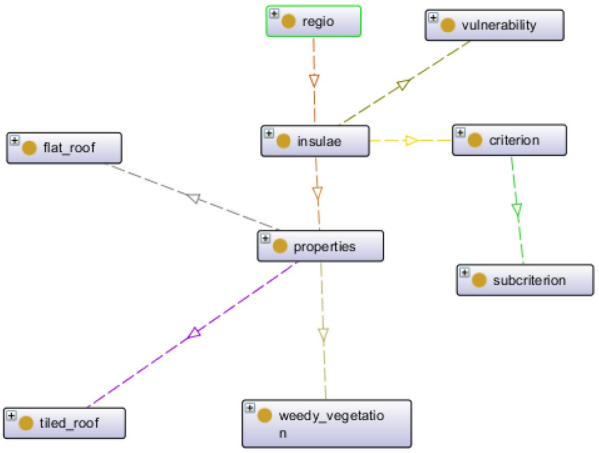


Figure 7.11: Pompeii Archaeological Site Ontology Graph

is composed of multiple structures, from which information such as name, location, and area size is derived.

As shown in Figure 7.12, when users hover over the CHVI label of an insula, they can view the information used to compute the static CHVI. This includes the state of use and various characteristics of the insula encoded in the underlying ontology, together with the number of maintenance issues detected within the area.

In Figure 7.13, pressing the structure information button allows users to display information about the structures by querying the underlying ontology of the site assets. When activated, labels corresponding to the structures appear on the map, enabling users to identify which structures are present in a given area and helping to mitigate the *memory trap* situation awareness demon.

7.3.3 VERGIL Adaptive Interface

The VERGIL map has a series of map display modes relating to maintenance issues that were identified during the GDTA phase, and in particular in sub-goals *1.1.1*, *1.1.2*, and *1.1.3*. Below, we will describe the various modes of this interface and how it is able to adapt to goals and risk situations in order to give suggestions to the user and encourage them to change goals when necessary, without being invasive.

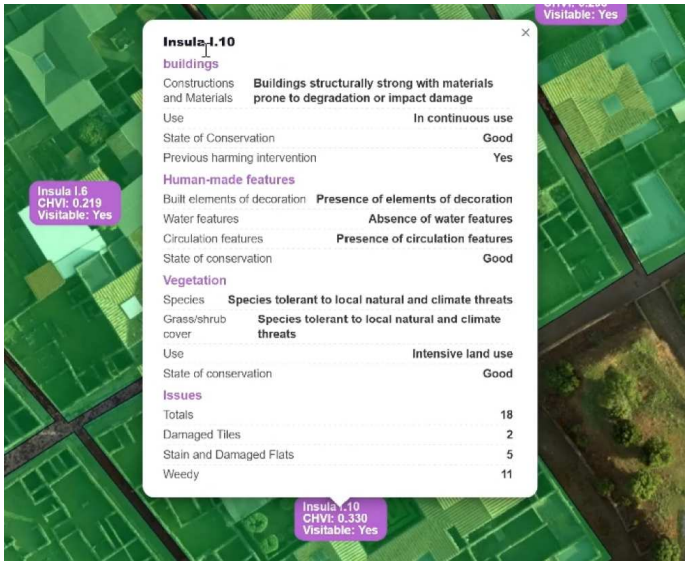


Figure 7.12: Vulnerability Info Hovering the Label of an Insula

7.3.3.1 Interface Modes

The modes of VERGIL are linked to the possibility of filtering specific maintenance issues so that human operators can inspect them depending on the user's current task.

The modes can be selected through a dedicated navbar, with four button related to each of the four modes:

- **Overview:** A mode that shows an overview of the condition of the site. Each insula has an associated vulnerability index that takes into account all maintenance issues and possible factors. When someone click on a specific insula to get details on maintenance issues, all the problems on that insula are shown, such as broken tiles, damaged conduits, and weedy vegetation. The $CHVI^d_{ov}$ is calculated as in equation 7.21;
- **Tiles:** The mode filters broken tiles on tiled roofs in the site, and the vulnerability index $CHVI^d_{ti}$ associated with the insula is calculated as in equation 7.24;
- **Flats:** The mode act as a filter on damaged and stained conduits on the flat roofs in the site, and the vulnerability index $CHVI^d_{fl}$ associated with the insula is calculated as in equation 7.26;
- **Weedy:** The mode is a related to weedy vegetation on both tiled and flat roofs, and the vulnerability index $CHVI^d_{wd}$ associated with the insula is

Figure 7.13: Label and Position of the Structures

calculated as in equation 7.28.

The mode is selected by the user according to their current goal. Since situation awareness is a goal-driven process, users decide which mode to use based on what they aim to accomplish. An example of mode selection is given in the Figure 7.14, in which the flat mode is selected.

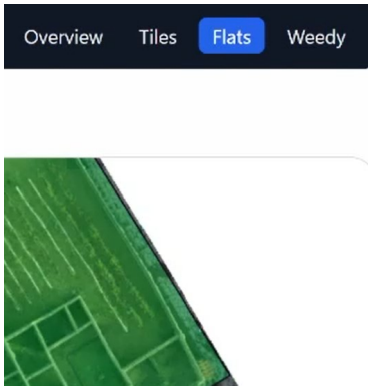


Figure 7.14: Flats Map of the SA Interface

7.3.3.2 Oscillatory Goal Suggestion Approach

The goal of the adaptive interface we propose in VERGIL is to address a fundamental issue in any situation awareness system: the natural trade-off between two complementary information-processing modes, namely goal-driven and data-driven processing. Situation awareness is inherently goal-driven be-

cause operators interpret the environment in terms of what they are trying to achieve [58]. Therefore, interface information should be organized around the user's current goal, rather than around data sources, so that the cues needed to decide and act are readily available and workload is reduced. However, a strong focus on the current goal can also amplify attentional tunneling, leading operators to concentrate on a narrow subset of information and miss relevant changes elsewhere. For this reason, situation awareness systems must also support global situation awareness, meaning a continuously accessible big picture of what is happening, so that operators can detect critical events, activate the appropriate mental models, and reprioritize goals when necessary. In our SA-CPSS, this process is handled by the projection module described in Section 4.2.3.4, namely **L3.2 Danger Projection**, introduced in Chapter 4, whose purpose is to identify potentially dangerous situations in the near future and draw attention to them when needed. This global monitoring is inherently data-driven because it must account for salient cues and anomalies that are not directly tied to the current goal.

Maintaining situation awareness therefore requires balancing these two modes: supporting the operator's goal without isolating attention, while redirecting attention toward higher-priority goals only when justified and without being disruptive or overly intrusive. If attention is captured at the wrong time, there is a risk of degrading situation awareness with respect to both the ongoing task and the emerging task the user is being redirected to. Alerts must therefore be gradual and appropriately adapted to the current situation, including their intensity. A similar issue arises when attempting to force a goal change, for example by automatically switching the interface mode. This can disorient the user, reduce comprehension, and ultimately degrade performance across all goals. For this reason, mode changes must remain under the user's control, allowing them to decide when to shift goals. In cases of high risk and imminent urgency, however, more intrusive alerts may be necessary to effectively capture attention.

To address these challenges, we propose an *Oscillatory Goal Suggestion* approach that alerts operators to the need for a goal switch. In this approach, the oscillation between goal-driven and data-driven processing is designed to resemble human attentional dynamics, making interaction more natural and less invasive. The method adopts an adaptive goal-suggestion mechanism that gradually shifts the operator's attention toward emerging priorities, mirroring how the brain transitions between states of alertness as early signs of an emergency appear [272, 273], rather than imposing abrupt switches that can cause confusion or loss of context . In this way, we aim to adapt the VERGIL interface in a more progressive and human-congruent manner, supporting a more dynamic and effective level of situation awareness across the overall system.

The approach we propose is grounded in neuroscientific evidence [272,273]

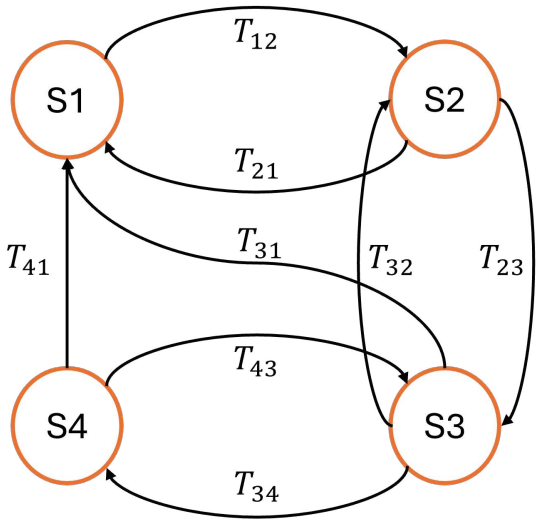


Figure 7.15: Alert States Model

showing that a user's attentional focus fluctuates across distinct cognitive states, each of which can be associated with specific ranges of brainwave frequencies [274]. These states are not merely theoretical constructs, but correspond to well-documented neurological conditions that reflect the brain's natural oscillation between goal-driven and data-driven processing. We therefore define the state space as follows:

- S_1. **Normal surveillance**: The system operates under standard conditions, performing routine monitoring mainly driven by historical data and baseline expectations. This state corresponds to low cognitive load and sustained attention, and can be associated with ϑ brain waves (4–8 Hz).

- S_2. **Moderate alert**: Triggered by early or weak signals of deviation from normal conditions, such as small but persistent increases in risk indicators. It can be associated primarily with α brain waves (8–13 Hz), reflecting relaxed but focused cognitive processing.

- S_3. **High alert**: Activated by more pronounced anomalies or rapidly evolving conditions that require heightened vigilance and faster decision-making. It can be associated with low-to-mid β brain waves (13–20 Hz).

- S_4. **Critical alert**: Represents an immediate response to highly exceptional and potentially dangerous events, such as extreme rainfall in the considered case study. It can be associated with high β brain waves (20–30 Hz).

The state model is shown in Figure 7.15, which also illustrates the possible transitions between states. Defining these transitions is critical. In our CHCI context, they can be specified through expert-based approaches, in which domain experts determine, based on variables such as rainfall intensity, cumulative rainfall, and rainfall duration, when a given goal requires more or less immediate attention. Alternatively, they can be learned through data-driven approaches.

7.3.3.3 VERGIL Goal Suggestion Implementation

In the VERGIL platform, we implement an adaptive interface in which transitions between states are driven by a set of thresholds based on rainfall intensity, cumulative rainfall, and rainfall duration. Indeed, these factors are highly dynamic and can quickly escalate into emergencies and cause serious damage to the park, especially when broken tiles are present. By contrast, more slowly varying factors such as temperature and humidity do not typically require this type of immediate attention, as their dynamics are slow. When these thresholds are exceeded, the system moves from one state to another. The thresholds are defined with the support of site experts, who specify which type of alert should be issued under given weather conditions when damaged tiles are present at the site.

The adaptive interface is also designed to draw attention to damaged tiles, when present, even while the user is operating in weedy or flat mode, by monitoring precipitation, which is the most critical factor in the presence of roof damage. This mechanism provides three levels of alert intensity, corresponding to states S_2, S_3, and S_4 in Figure 7.15, and includes state S_1, in which no attention shift is required. Every change in status must also be communicated to the user, which in our case is done through various visual signals.

The states are described as follows:

- **No alert:** The user performs the task normally, and no alerts are active because the system has not detected any potential risk situations. The interface appears as shown in Figure 7.7. This corresponds to state S_1 in Figure 7.15;
- **Moderate alert:** a fixed yellow alert appears on the broken-tiles button in the navbar, indicating that it may be advisable to switch the goal to identifying broken tiles as soon as possible. Figure 7.16 shows the S_2 yellow alert on the interface;
- **High alert:** a blinking orange alert appears on the broken-tiles button in the navbar, indicating that the risk of serious damage is becoming imminent. Figure 7.17 shows the S_3 orange alert on the interface;

- **Critical alert:** a blinking red alert appears on the broken-tiles button in the navbar, indicating that the risk is serious and imminent and requires a mode change. Given the urgency, a pop-up also appears at the bottom of the interface, prompting the user to switch. Figure 7.18 shows the S_4 red alert on the interface.

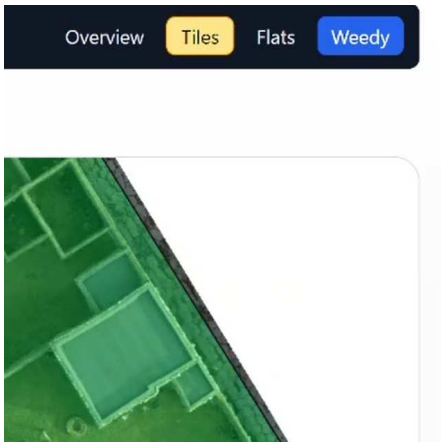


Figure 7.16: Yellow Alert for Rainfall

Figure 7.17: Orange Alert for Rainfall

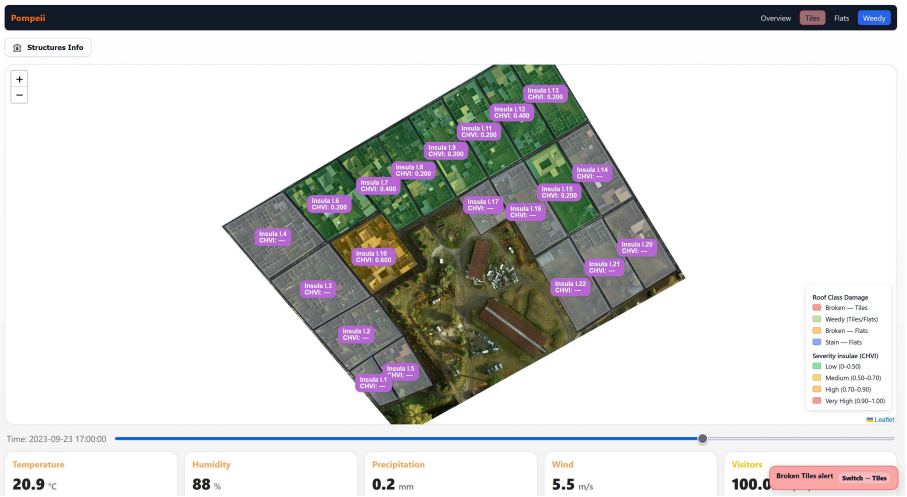


Figure 7.18: Red Alert for Rainfall

Below we formally and rigorously define the model we propose:

Let $t \in \mathbb{Z}_{\geq 0}$ denote the hourly time index. R_t is the hourly rainfall (mm/h) at time t: a detail worth noting is that this value does not reflect only the current

conditions, but is computed as the maximum expected value over the eight hours following time t, so that the user is warned in advance and has time to take precautions if the near-term forecast is adverse; R_t^c, instead, is the exponentially weighted cumulative rainfall indicator at time t, while $A_t \in \{\mathrm{S}_1, \mathrm{S}_2, \mathrm{S}_3, \mathrm{S}_4\}$ is the set of states of the interface (No alert, Moderate Alert, High Alert, Critical Alert).

The transition from one alert level to another is regulated by a series of thresholds connected to various variables. The first variable to consider is the hourly rainfall R_t, for which the threshold values are:

$$TH^R_{\mathrm{S}_2} = 2, \qquad TH^R_{\mathrm{S}_3} = 7, \qquad TH^R_{\mathrm{S}_4} = 20, \tag{7.1}$$

where $TH^R_{\mathrm{S}_2}$ is the rainfall threshold above which the system switches to state S_2, $TH^R_{\mathrm{S}_3}$ is the threshold for state S_3, and $TH^R_{\mathrm{S}_4}$ is the threshold for state S_4. When R_t exceeds one of these thresholds, the system transitions to the corresponding alert state.

Then there is the cumulative rainfall R_t^c

$$TH^{R^c}_{\mathrm{S}_2} = 5, \qquad TH^{R^c}_{\mathrm{S}_3} = 12, \qquad TH^{R^c}_{\mathrm{S}_4} = 25, \tag{7.2}$$

where $TH^{R^c}_{\mathrm{S}_2}$ is the cumulative rainfall threshold above which the system switches to state S_2, $TH^{R^c}_{\mathrm{S}_3}$ is the threshold for state S_3, and $TH^{R^c}_{\mathrm{S}_4}$ is the threshold for state S_4. Similarly, when the cumulative rainfall R_t^c exceeds one of these thresholds, the system transitions to the corresponding alert state.

Lastly, we consider persistence of rain variable h_t which counts how many consecutive hours the cumulative indicator has remained at least moderate ($R^c \geq TH^{R^c}_{\mathrm{S}_2}$):

$$\begin{aligned} h_t = \max \Big\{ h \in \mathbb{N} :\ & R^c_{t-k} \geq TH^{R^c}_{\mathrm{S}_2} \ \ \forall k = 0, \ldots, h-1 \Big\}, \\ & (h_t = 0 \text{ if } R^c_t < TH^{R^c}_{\mathrm{S}_2}). \end{aligned} \tag{7.3}$$

For such values the relative threshold are:

$$TH^h_{\mathrm{S}_2} = 1, \qquad TH^h_{\mathrm{S}_3} = 2, \qquad TH^h_{\mathrm{S}_4} = 4. \tag{7.4}$$

where $TH^h_{\mathrm{S}_2}$ is the rainfall-duration threshold above which the system switches to state S_2, $TH^h_{\mathrm{S}_3}$ is the threshold for state S_3, and $TH^h_{\mathrm{S}_4}$ is the threshold for state S_4. When h_t exceeds one of these thresholds, the system transitions to the corresponding alert state.

Each alert state is persistent until the goal, and the relative mode interface, is not changed. The logic for transitioning into each alert status is formally defined as follows:

- $\boldsymbol{S_4}$**- Critical alert state:** it is indicated by the red blinking alert and the popup, and it is triggered either by extreme hourly rain, or by long persistence ($h_t \geq TH^h_{\mathrm{S}_4}$). The boolean condition $\mathcal{R}_t$ is reported in the following:

$$\mathcal{R}_t := (A_{t-1} = \mathrm{S}_4) \vee (R_t \geq TH^R_{\mathrm{S}_4}) \vee (R^c_t \geq TH^{R^c}_{\mathrm{S}_4}) \vee (h_t \geq TH^h_{\mathrm{S}_4}). \tag{7.5}$$

- $\boldsymbol{S_3}$**- High alert state:** If not S_4, S_3 state, which is associated to the orange blinking alert, is triggered by high hourly rain, by high cumulative indicator, by moderate persistence. The boolean condition $\mathcal{O}_t$ is reported in the following:

$$\mathcal{O}_t \ := \ \big(A_{t-1} = \mathrm{S}_3\big) \vee \big(R_t \geq TH^R_{\mathrm{S}_3}\big) \vee \big(R^c_t \geq TH^{R^c}_{\mathrm{S}_3}\big) \vee \big(h_t \geq TH^h_{\mathrm{S}_3}\big). \tag{7.6}$$

- $\boldsymbol{S_2}$**- Moderate alert state:** If neither S_4 nor S_3, S_2 state, which is associated to the yellow static alert, is triggered by moderate hourly rain, by moderate cumulative indicator. The boolean condition $\mathcal{Y}_t$ is reported in the following:

$$\mathcal{Y}_t \ := \ \big(A_{t-1} = \mathrm{S}_2\big) \vee \big(R_t \geq TH^R_{\mathrm{S}_2}\big) \vee \big(R^c_t \geq TH^{R^c}_{\mathrm{S}_2}\big) \vee \big(h_t \geq TH^h_{\mathrm{S}_2}\big). \tag{7.7}$$

The alert level is updated with strict priority $\mathrm{S}_4 > \mathrm{S}_3 > \mathrm{S}_2 > \mathrm{S}_1$:

$$A_t = \begin{cases} \mathrm{S}_4 & \text{if } \mathcal{R}_t, \\ \mathrm{S}_3 & \text{if } \neg\mathcal{R}_t \wedge \mathcal{O}_t, \\ \mathrm{S}_2 & \text{if } \neg\mathcal{R}_t \wedge \neg\mathcal{O}_t \wedge \mathcal{Y}_t, \\ \mathrm{S}_1 & \text{otherwise.} \end{cases} \tag{7.8}$$

Through Equation 7.8, transitions between alert states are triggered while the system is in flat or weedy mode, so that the user is guided toward tile mode whenever maintenance issues related to damaged tiles are present. This threshold-based approach implements the model described in Section 7.3.3.2 and the projection module presented in Section 4.2.3.4 of the proposed SA-CPSS. In particular, as described in the *L3.2 Danger Projection module*, it enables users to respond to emergency situations more promptly than relying on the current CHVI alone, and supports proactive anticipation of near-term risks.

7.3.4 CHVIs Implementation

In Section 4.3, we introduced a vulnerability measure that was then used in Chapter 4 and applied again during the evaluation phase, in the implementation described in Section 4.4.1.

However, as mentioned above, the CHVI we designed is intended as a goal-driven vulnerability measure, whose computation must be performed according to the user's current goal. In VERGIL, the human operator pursues their goal by selecting the appropriate mode. We therefore present the implementation of four vulnerability indices, one for each mode of the platform.

Since the contributions introduced in Section 4.4.1.2 were illustrative examples used for our proof of concept, they have been revised more rigorously together with domain experts. These experts supported us in defining the thresholds according to the specific area under analysis and the domain covered by our experimentation.

7.3.4.1 Contribution Functions

Since the role of these variables in assessing the vulnerability of the areas has been discussed in Section 4.4.1.2, this section presents the updated contributions and discusses the differences from a goal-driven perspective.

The symbols used are the same as those introduced in Section 4.4.1.2, and are repeated here for the reader's convenience. We consider three types of maintenance issues, namely damaged conduits (x_{dc}), broken tiles (x_{bt}), and weedy vegetation (x_{wv}). We also consider meteorological variables including precipitation (x_p), temperature (x_t), relative humidity (x_u), and wind speed (x_w), as well as visitor flows (x_v). The resulting context space is therefore $X = \{x_{dc}, x_{bt}, x_{wv}, x_p, x_t, x_u, x_w, x_v\}$, and a contribution function must be defined for each dimension.

Equations 7.9, 7.10, and 7.11 present the updated contributions, reflecting the fact that the areas under consideration are larger and denser in terms of structures and, consequently, more affected by maintenance issues. Moreover, these contributions have been adapted to a new approach for issue identification, which no longer relies on detection but instead on classification applied to pre-segmented areas during the initial setup phase.

Equations 7.12, 7.14, 7.15, and 7.17 introduce new contributions in the meteorological domain, while the equation for visitors (7.19) remains unchanged with respect to Section 4.4.1.2.

Finally, Equations 7.13, 7.16, 7.18, and 7.20 define the parameters used to adapt contextual contributions to the detected maintenance issues. As can be observed, these parameters must be differentiated according to the user's

goal, resulting in contribution functions that vary not only with the detected problems but also with the user's current objective. Furthermore, while Chapter 4 addressed a proof of concept in which the maximum number of damages was assumed to be fixed based on the scenario and the structures analyzed, the new formulation generalizes the parameters without imposing an upper limit for the damages in an Insulae.

$$con(x_{bt}) = \begin{cases} 0 & x_{bt} = 0, \\ 0.2 & x_{bt} \in (0, 1], \\ 0.4 & x_{bt} \in (1, 2], \\ 0.5 & x_{bt} \in (2, 3], \\ 0.6 & x_{bt} \in (3, 5], \\ 0.7 & x_{bt} \in (5, 10], \\ 0.8 & x_{bt} \in (10, 20], \\ 0.9 & x_{bt} \in (20, 30], \\ 1 & x_{bt} \in (30, +\inf]. \end{cases} \tag{7.9}$$

$$con(x_{dc}) = \begin{cases} 0 & x_{dc} = 0, \\ 0.1 & x_{dc} \in (0, 1], \\ 0.2 & x_{dc} \in (1, 3], \\ 0.3 & x_{dc} \in (3, 5], \\ 0.5 & x_{dc} \in (5, 7], \\ 0.7 & x_{dc} \in (7, 15], \\ 0.8 & x_{dc} \in (15, 25], \\ 0.9 & x_{dc} \in (25, 30], \\ 1 & x_{dc} \in (30, +\inf]. \end{cases} \tag{7.10}$$

$$con(x_{wv}) = \begin{cases} 0 & x_{wv} = 0, \\ 0.1 & x_{wv} \in (0, 1], \\ 0.2 & x_{wv} \in (1, 4], \\ 0.3 & x_{wv} \in (4, 6], \\ 0.5 & x_{wv} \in (6, 8], \\ 0.6 & x_{wv} \in (8, 10], \\ 0.8 & x_{wv} \in (10, 20], \\ 1 & x_{wv} \in (20, +\inf]. \end{cases} \tag{7.11}$$

$$con(x_p) = \begin{cases} 0 & x_p = 0, \\ 0.2 & x_p \in (0, 2.5 - 2\rho], \\ 0.5 & x_p \in (2.5 - 2\rho, 10 - 5\rho], \\ 0.8 & x_p \in (10 - 5\rho, 50 - 20\rho], \\ 1 & x_p \in (50 - 20\rho, +\inf]. \end{cases} \tag{7.12}$$

$$\rho = \begin{cases} (0.7\dfrac{x_{bt}}{x_{bt} + \alpha} + 0.3\dfrac{x_{dc}}{x_{dc} + \beta}) & \text{if mode} = \textit{Overview}, \\ \dfrac{x_{bt}}{x_{bt} + \alpha} & \text{if mode} = \textit{Tiles}, \\ \dfrac{x_{dc}}{x_{dc} + \beta} & \text{if mode} = \textit{Flats}, \\ 1 & \text{otherwise.} \end{cases} \tag{7.13}$$

$$con(x_t) = \begin{cases} 0.1 & x_t \in (-\inf, -5], \\ 0.2 & x_t \in (-5, 0], \\ 0.1 & x_t \in (0, 10], \\ 0.0 & x_t \in (10, 28], \\ 0.2 & x_t \in (28, 32], \\ 0.3 & x_t \in (32, 35], \\ 0.4 & x_t \in (35, 37], \\ 0.5 & x_t \in (37, 42], \\ 1.0 & x_t \in (42, +\inf]. \end{cases} \tag{7.14}$$

$$con(x_u) = \begin{cases} 0.4 & x_u \in [0, 20), \\ 0.2 & x_u \in [20, 40), \\ 0.0 & x_u \in [40, 60), \\ 0.6 & x_u \in [60, 80 - 5\mu), \\ 0.8 & x_u \in [80 - 5\mu, 90 - 5\mu), \\ 1.0 & x_u \in [90 - 5\mu, 100). \end{cases} \tag{7.15}$$

$$\mu = \begin{cases} \dfrac{x_{dc}}{x_{dc} + \beta} & \text{if mode} = \textit{Flats}, \\ 1 & \text{otherwise.} \end{cases} \tag{7.16}$$

$$
con(x_w) = \begin{cases}
0.0 & x_w \in [0, 5.4], \\
0.1 & x_w \in (5.4, 7.9 - 2\omega], \\
0.2 & x_w \in (7.9 - 2\omega, 13.8 - 4\omega], \\
0.5 & x_w \in (13.8 - 4\omega, 17.1 - 6\omega], \\
0.6 & x_w \in (17.1 - 6\omega, 20.7 - 8\omega], \\
0.7 & x_w \in (20.7 - 8\omega, 24.4 - 12\omega], \\
1.0 & x_w \in (24.4 - 12\omega, +\inf].
\end{cases} \tag{7.17}
$$

$$
\omega = \begin{cases}
(0.5\dfrac{x_{bt}}{x_{bt} + \alpha} + 0.5\dfrac{x_{dc}}{x_{dc} + \beta}) & \text{if mode} = \textit{Overview}, \\
\dfrac{x_{bt}}{x_{bt} + \alpha} & \text{if mode} = \textit{Tiles}, \\
\dfrac{x_{dc}}{x_{dc} + \beta} & \text{if mode} = \textit{Flats}, \\
1 & \text{otherwise.}
\end{cases} \tag{7.18}
$$

$$
con(x_v) = \begin{cases}
0 & x_v = 0, \\
0.1 & x_v \in (0, 12500 - 5000\lambda], \\
0.2 & x_v \in (12500 - 5000\lambda, 25000 - 5000\lambda], \\
0.3 & x_v \in (25000 - 5000\lambda, 37000 - 10000\lambda], \\
0.4 & x_v \in (37000 - 10000\lambda, 50000 - 20000\lambda], \\
0.5 & x_v \in (50000 - 20000\lambda, 62500 - 30000\lambda], \\
0.6 & x_v \in (62500 - 30000\lambda, 75000 - 40000\lambda], \\
0.7 & x_v \in (75000 - 40000\lambda, 87500 - 50000\lambda], \\
0.8 & x_v \in (87500 - 50000\lambda, 100000 - 60000\lambda], \\
0.9 & x_v \in (100000 - 60000\lambda, 112500 - 70000\lambda], \\
1 & x_v \in (112500 - 70000\lambda, +\inf].
\end{cases} \tag{7.19}
$$

$$
\lambda = \begin{cases}
\dfrac{x_{bt}}{x_{bt} + \alpha} & \text{if mode} = \textit{Tiles}, \\
1 & \text{otherwise.}
\end{cases} \tag{7.20}
$$

The parameters α and β were determined from the median damage counts for broken tiles and damaged conduits across all insulae exhibiting damage. This yielded $\alpha = 1.5$ and $\beta = 2$.

7.3.4.2 *Goal-driven CHVIs*

After introducing the new contribution functions, we define the implementation of each CHVI associated with the various areas for each of the four modes described above.

The first is the $CHVI_{ov}$, which corresponds to the overview mode. As described in Section 4.4.1, this CHVI is composed of two components: a static component, denoted as $CHVI^s_{ov}$, which is based on materials, usage, and other static contributions defined by domain experts, and a dynamic component, denoted as $CHVI^d_{ov}$, which is computed using CST.

$$CHVI_{ov} = w^s_{ov} \cdot CHVI^s_{ov} + w^d_{ov} \cdot CHVI^d_{ov} \tag{7.21}$$

with weights $w^s_{ov} = 0.25$ and $w^d_{ov} = 0.75$, which result in the dynamic component having a slightly greater influence in the final computation. While $CHVI^s_{ov}$ is calculated as described in Section 4.4.1.1, $CHVI^d_{ov}$ is computed according to Equation 7.22:

$$\begin{aligned} CHVI^d_{ov} = w_{ov,bt}\,\mathrm{con}(x_{bt}) + w_{ov,dc}\,\mathrm{con}(x_{dc}) + w_{ov,wv}\,\mathrm{con}(x_{wv}) + \\ w_{ov,p}\,\mathrm{con}(x_p) + w_{ov,t}\,\mathrm{con}(x_t) + w_{ov,u}\,\mathrm{con}(x_u) + w_{ov,w}\,\mathrm{con}(x_w) + \\ w_{ov,v}\,\mathrm{con}(x_v) \end{aligned} \tag{7.22}$$

with the weights assuming the following values:

$$\begin{aligned} w_{ov,bt} = 0.40, \quad w_{ov,dc} = 0.19, \quad w_{ov,wv} = 0.05, \quad w_{ov,p} = 0.20, \\ w_{ov,t} = 0.0025, \quad w_{ov,u} = 0.0075, \quad w_{ov,w} = 0.05, \quad w_{ov,v} = 0.10 \end{aligned} \tag{7.23}$$

In our implementation, the other three indices associated with the remaining modes, namely tiles, flats, and weedy, rely exclusively on the dynamic component. In this way, the *insulae* that are not affected by the issues filtered by a given mode are excluded from the calculation, allowing the operator to focus more directly on actual maintenance problems with a real-time impact. With this approach, such impact is not constrained by the static component, which, during a phase focused on a specific goal, could otherwise be unnecessarily distracting.

The first CHVI we analyse is the one related to the tiles mode, denoted as $CHVI^d_{ti}$, in which the contributions are provided by damaged tiles, as well as by rain, wind, and visitors:

$$CHVI^d_{ti} = w_{ti,bt}\,\mathrm{con}(x_{bt}) + w_{ti,p}\,\mathrm{con}(x_p) + w_{ti,w}\,\mathrm{con}(x_w) + w_{ti,v}\,\mathrm{con}(x_v) \tag{7.24}$$

with the contributions weighted so as to assign greater importance to broken tiles and rain, and lower contribution to wind and visitors.

$$w_{ti,bt} = 0.50, \quad w_{ti,p} = 0.30, \quad w_{ti,w} = 0.10, \quad w_{ti,v} = 0.10 \tag{7.25}$$

Then, we consider the index related to the flats mode, denoted as $CHVI^d_{fl}$, in which the contributions are provided by damaged conduits, as well as by rain, wind, and visitors:

$$CHVI^d_{fl} = w_{fl,dc}\,\mathrm{con}(x_{dc}) + w_{fl,p}\,\mathrm{con}(x_p) + w_{fl,w}\,\mathrm{con}(x_w) + w_{fl,v}\,\mathrm{con}(x_v) \tag{7.26}$$

In this case, a greater weight is assigned to maintenance issues, while rain and wind contribute equally with respect to each other, since wind can be particularly critical in the presence of broken conduits, together with a lower contribution from visitors.

$$w_{fl,dc} = 0.50, \quad w_{fl,p} = 0.20, \quad w_{fl,w} = 0.20, \quad w_{fl,v} = 0.10 \tag{7.27}$$

Lastly, the CHVI related to weedy vegetation, denoted as $CHVI^d_{wd}$, takes into account only the contribution of weedy vegetation issues and visitors, with equal weighting:

$$CHVI^d_{wd} = w_{wd,wv}\,\mathrm{con}(x_{wv}) + w_{wd,v}\,\mathrm{con}(x_v) \tag{7.28}$$

since the other contextual components do not have a significant impact on the level of danger posed by this maintenance issue.

$$w_{wd,wv} = 0.50, \quad w_{wd,v} = 0.50 \tag{7.29}$$

Each of these indices comes into play at the appropriate time, and the type of CHVI therefore changes whenever the user's goal changes, since it is the user who determines the displayed information.

In the next section, we evaluate the extent to which these indices can support operators in decision-making processes, both in routine activities and in emergency situations.

7.4 Evaluation

In this section, we evaluate the SA-CPSS methodological approach for cultural heritage protection in its various aspects using the SAGAT method and the

VERGIL platform deployed at the archaeological site of Pompeii. We compare VERGIL with a baseline system that includes some of its features, such as the automatic identification of maintenance issues using HydraML (6.3), but does not implement several key elements of our CPSS. In particular, the baseline lacks part of the understanding phase and all projection modules. Indeed, it does not provide any visualization of area vulnerability through CHVI indices for the insulae of Regio I, as it does not explicitly implement a situational model of the kind adopted in our SA-CPSS. In addition, the Oscillatory Goal Suggestion Approach for emergency alerts is implemented only in VERGIL.

Participants will test both the baseline and VERGIL in a set of usage scenarios with predefined objectives. Through questionnaires, we will assess whether VERGIL leads to measurable improvements, which aspects benefit the most, and, crucially, whether VERGIL users are able to make better decisions than when using the baseline system. Overall, these results will indicate whether the features introduced by our SA-CPSS provide tangible benefits.

7.4.1 Evaluation Procedure

In order to truly measure the improvement in terms of SA linked to the use of the system we propose, we must involve human users since, as Endsley said, *"the true situation awareness only exists in the mind of the human operator"* [59]. For this reason, our experiment involved 24 human subjects with heterogeneous characteristics, all sharing partial or complete knowledge of the domain of interest. The participants included computer engineers with experience in SA support across different domains, computer scientists and electronic engineers working in the field of CH, civil engineers with expertise in CH, and specific domain experts.

Each subject was tested in five different usage scenarios, each referring to a time window of about 18 hours between September 22 and 23, 2023. The simulations are run with an accelerated data update rate, so that one hour of simulated time corresponds to one minute of real-time, and the time slider advances accordingly. By usage scenarios, we mean *"a formally specified, time-ordered instantiation of an operational context (mission/task) that defines the evolving "world state" (system + environment), the operator's goals and tasks, and the event timeline, such that the operator's SA can be sampled at selected instants and scored against an objective ground truth"* [21].

In general, however, SA is assessed by comparing a system against an alternative, because the key question is whether the innovations introduced actually improve the baseline with respect to the previously elicited GDTA, because through that we understood what the operator must comprehend and

Figure 7.19: Overview Map of the Base Interface

decide. In our case, the baseline shown in Figure 7.19 represents a system that already implements many of the features described above, including the identification of maintenance issues using HydraML, but it does not directly model the concept of situation as a measure of area vulnerability and it does not implement an the Oscillatory Goal Suggestion Approach to support goal change in emergency situations. By comparing this system, we will therefore be able to evaluate the contribution in terms of these aspects, which are at the heart of the SA-based approach proposed with our CPSS. Indeed, in the baseline system, the operator can view maintenance issues and contextual data through the different map widgets, but there are no indicators that express area vulnerability as the combination of these factors, and no visual cues that allow the operator to assess the status of the insulae as a fusion of this information. In addition, the baseline interface is not adaptive to the situation. The system cannot alert the user that something is changing, or that it may be necessary to update goals and adjust the visualization accordingly.

With this in mind, each subject followed the procedure below. First, each participant was trained through a VERGIL demo that introduced the system's main features and explained what to focus on during the different usage scenarios in the archaeological site of Pompeii, set between September 22 and 23, 2023, and based on real information and data from that period as presented by the interface. Next, the participant was tested on five scenarios. For each scenario, it was pseudorandomly determined whether the participant would use the baseline system or VERGIL, with the users not knowing which of the two they are using. The assignment was balanced so that, for every scenario, half of the users worked with one system and half with the other, resulting in three scenarios with one system and two with the other, allowing each operator to become familiar with both. The order of the scenarios was also randomized

across participants. After each scenario, a questionnaire was administered with questions targeting perception, comprehension, and projection. Each answer was scored up to one point, with intermediate scores assigned to partially correct responses.

The collected data were then interpreted in several ways. Outliers were removed when they were clearly attributable to users having little or no understanding of a given scenario, or to answers that appeared to be based on recollections from other scenarios. For each questionnaire, and therefore for each user and scenario, a score is calculated, in which each question is assigned a maximum of 1 point for a correct answer, with the possibility of intermediate scores between 0 and 1 if the user has answered partially correctly. For example, if the question is: which insula would you intervene on first in the event of broken tiles, the user gets 1 if they identify the insula with the highest maintenance priority, 0.5 or more if they identify an insula that is also at risk but not a top priority, or zero points if they identify irrelevant insulae.

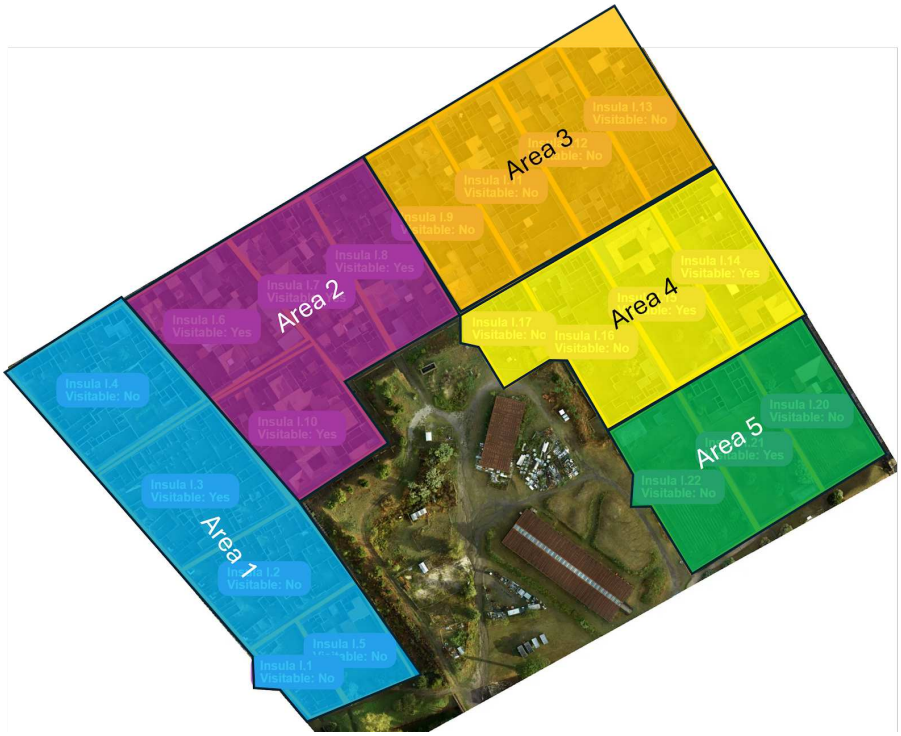


Figure 7.20: Areas of Regio I

We computed the mean and standard deviation of the questionnaire scores for users on both systems, baseline and VERGIL. If the mean score is higher for VERGIL, this can indicate an improvement in performance enabled by

our system. We repeated the same analysis for perception, comprehension, and projection separately, in order to assess improvements at each level. This comparison supports an evaluation of overall improvements in terms of SA. We also analysed overall SA and the three levels on a per-scenario basis, to better understand where the improvements emerged.

Since differences in mean and standard deviation alone do not indicate whether the observed improvements are due to chance, we also performed statistical significance tests.

Figure 7.21: Insulae of Regio I

7.4.2 Evaluation Scenarios

At this point, it is important to define the usage scenarios designed to evaluate our system. We tested five scenarios. Each participant completed all five, but used only one of the two systems, Baseline or VERGIL.

At the start of each test, we introduced the scenario's main objective and provided brief instructions on what to focus on, especially for participants with less experience in the domain.

Most scenarios lasted only a few minutes, around 3 to 3.30. At the end

of each one, participants completed a questionnaire. The final scenario lasted about 9 minutes and was divided into four segments. After each segment, participants answered a short set of questions.

In the following sections, we describe each scenario and list the questions that were asked. Not all questions were included in the final analysis. Some items were open-ended and were used to assess participants' understanding of the scenario, helping identify outliers, while others were included as distractors.

When describing the scenarios, we will refer to the area and insulae numbering shown in Figures 7.20 and 7.21, respectively.

7.4.2.1 Scenario 1

In the first scenario, the user must determine which areas most urgently require pruning of weedy vegetation on roofs. The relative questionnaire is in Table 7.2. There are six questions; therefore, the maximum score a user can achieve is 6. The user identifies the most critical areas to address first by flying over the map. At this stage, the goal is not to inspect individual areas in detail, but to gain a broad overview and decide where to focus. The simulation runs from 08:00 p.m. on September 22, 2023, to 11:00 p.m. on the same day.

The user first selects the overview mode and then switches to the weedy vegetation mode to examine the related maintenance issues in more detail. Based on the overview mode, the user is also expected to assess whether any areas may require attention in the event of future rainfall.

During the task, the user may notice maintenance issues and vulnerabilities that are not only linked to weedy vegetation, but also related to broader roof conditions, such as damaged roofs or broken tiles. They are then asked questions to verify whether they identified areas that could require attention in case of rain, even when this is not directly tied to weedy vegetation. For instance, the first projection question in Table 7.2 targets this aspect. The highest score is awarded when the user identifies the insula where roof repairs are most urgent. Lower scores are assigned when the reference to specific insulae is vague or confused, or when the user points to insulae where repairs are needed but with lower urgency. In this scenario, the highest priority insula was number 15.

The second question in the perception phase is also intended to assess whether the user noticed that Insula 15 is vulnerable due to roof damage and may therefore be a priority. However, if the user does not identify this and instead focuses only on weedy vegetation and on the insulae that are most critical from that perspective, such as Insula 10, the answer is scored as partially correct. The maximum score is assigned when the user correctly identifies the individual insulae and can keep both priorities in mind, namely weed infestation and the risk associated with broken tiles. In this scenario, the most accurate

SA level	Question type	Question
Perception		
	Multiple-choice	Was it raining during the simulation?
	Multiple-choice	Do the forecasts say it will rain in the next few hours?
Comprehension		
	Multiple-response	Looking at the figure above, which areas of the map urgently require the removal of invasive vegetation?
	Open-ended	Can you identify the insula(e) where this removal is most urgent?
Projection		
	Open-ended	In the event of rain, which insula(e) would you secure first?
	Open-ended	Based on what you saw in the video, which areas or insulae would you like to examine in more detail with a zoom?

Table 7.2: Scenario 1 Questionnaire.

answers referred to Area 2, Insula 10, and to Insula 15 in Area 4.

The comprehension questions, instead, focus on assessing SA related to weedy vegetation. Areas 2 and 4 have the highest priority for pruning, with Insula 10 as the top priority, followed by Insula 7. We also evaluate whether the user can correctly recognize this order of priority, since it supports the definition of the maintenance plan.

Perception level questions are also included to assess whether the user noticed ongoing or imminent rainfall, based on the forecasts.

7.4.2.2 Scenario 2

In the second scenario, the user must determine which areas most urgently require the replacement of damaged or deteriorated conduit on flat roofs. The relative questionnaire is in Table 7.3. There are five questions; therefore, the maximum score a user can achieve is 5. The user identifies the most critical areas to address first by flying over the map. At this stage, the goal is not to inspect individual areas in detail, but to gain a broad overview and decide where to focus. The simulation runs from 11:00 p.m. on September 22, 2023, to 1:00 a.m. on the following day.

The user first selects the flats mode and, at a certain point, switches to the tiles mode. In this mode, the user can identify the areas that are most exposed to heavy rainfall.

The questions, divided by SA level, are shown in Table 7.3. Based on the

SA level	Question type	Question
Perception		
	Multiple-choice	Did you notice wind during the simulation?
Comprehension		
	Multiple-response	In which areas was there the highest concentration of flat roofs with damaged waterproof membranes (damage or moisture stains)?
	Open-ended	Can you identify the insula(e) where replacing the waterproof membranes is most urgent?
Projection		
	Open-ended	Based on what you saw in the video, in the event of rain, would you add any insulae to the previous ones among those requiring imminent maintenance?
	Open-ended	Based on what you saw in the video, which areas or insulae would you like to see in greater detail with a zoom?

Table 7.3: Scenario 2 Questionnaire.

scenario, the user should recognize that the insulae where damaged sheathing requires the most urgent intervention are Insula 9 and Insula 16.

However, Insula 15, which shows extensive tile damage and deterioration, is the most susceptible to severe consequences in the event of rain. It should therefore be addressed even if it is not the user's primary goal at that moment.

The final projection question is meant to assess how many of these aspects the user noticed and, most importantly, whether the user can balance multiple maintenance priorities. As in the other scenarios, scoring is progressive and depends on the level of detail the user provides, including whether they recall specific areas or insulae, and whether they correctly identify the highest priority ones.

A perception level question related to the observed wind is also included.

7.4.2.3 Scenario 3

In this scenario, the user's goal is to check area 3 and 4. The user can inspect the insulae within these areas and must determine which ones require the most urgent intervention for conduit replacement. The relative questionnaire is in Table 7.4. There are seven questions; therefore, the maximum score a user can achieve is 7. The simulation runs from 12:00 a.m. on September 23, 2023, to 2:00 a.m. on the same day.

The comprehension questions are intended to assess whether the user recognizes that the most critical insulae are 9 and 16, and whether they correctly

identify Insula 9 as the one that becomes critical first, and therefore requires immediate intervention.

SA level	Question type	Question
Perception		
	Multiple-choice	Were these insulae accessible to visitors?
	Multiple-choice	Was the humidity at levels that could contribute to the deterioration of stains on the waterproof membranes and on the structures' walls?
	Multiple-choice	Was there wind during the video?
Comprehension		
	Multiple-response	Which insulae most urgently require replacement of damaged waterproof membranes on the flat roofs?
	Multiple-choice	Which of these would you have addressed first?
Projection		
	Multiple-response	Based on what you saw in the video, in the event of rain, which insula(e) should be given priority for maintenance operations?
	Multiple-choice	Will the wind increase, decrease, or remain stable?

Table 7.4: Scenario 3 Questionnaire.

From a projection perspective, it is also important to determine which areas are most vulnerable in the event of rain. At a certain point, the weather forecast starts indicating imminent rainfall, potentially even heavy rain. As a result, the user's goal shifts, and checking for broken tiles becomes relevant. Alongside Insulae 9 and 16, Insula 15 must also be considered, since it is clearly affected by broken tiles in tiles mode. In terms of vulnerability, Insulae 10 and 8 also represent a risk, although less severe than Insula 15.

This scenario also includes several perception questions, focusing on the accessibility of the insulae, as well as observed and forecast wind conditions and humidity. These questions are reported in Table 7.4.

7.4.2.4 Scenario 4

In this scenario, the user's goal is to inspect the Area 2. The user must inspect the four insulae within this area in order to determine which one should be addressed first in terms of weedy vegetation. The relative questionnaire is in Table 7.5. The simulation runs from 12:00 a.m. on September 23, 2023, to 3:00 a.m. on the same day. The questions for this scenario are reported in Table 7.5. There are seven questions; therefore, the maximum score a user can achieve is 7.

SA level	Question type	Question
Perception		
	Multiple-choice	Did it rain during the video?
	Multiple-choice	Do the forecasts say it will rain in the next few hours?
Comprehension		
	Multiple-choice	Which insula should be given absolute priority for invasive vegetation maintenance, given that it was in the worst condition?
	Multiple-choice	Based on what you saw across the various insulae, do you think it is easy to carry out combined interventions on both invasive vegetation and broken roof tiles?
	Open-ended	If yes, on which insulae can these combined interventions be carried out?
Projection		
	Multiple-response	In the event of rain, which areas of Pompeii should be monitored and checked first for broken roof tiles?
	Open-ended	In particular, which insula do you expect to be most damaged by the rain?

Table 7.5: Scenario 4 Questionnaire.

The operator must identify the insula with the highest priority for pruning, and also determine which insula has priority from a rainfall perspective. Toward the end of the video, the user's goal shifts to checking for broken tiles, and the display is updated accordingly. The user is then asked which areas are most at risk in case of rain.

In this case, Insula 15 has the highest priority for rain related maintenance, while Insulae 10 and 8 follow with lower priority. The latter are also the insulae where combined interventions can be planned, since they are vulnerable from both perspectives, namely weedy vegetation and broken tiles.

Users are also asked perception questions about rainfall, both observed and forecast, to assess whether they noticed rain that has already occurred or is imminent.

7.4.2.5 Scenario 5

Scenario 5 is the longest of the usage scenarios in this experiment. The relative questionnaire is in Table 7.6. There are twentytwo questions; therefore, the maximum score a user can achieve is 22. The simulation runs from 2:00 a.m. on September 23, 2023, to 11:00 a.m. on the same day, for a total duration of 9 minutes.

A scenario of this length cannot be used without adjustments, mainly be-

cause it would be difficult to ask questions about the early part of the video without placing an excessive load on the user's memory. For this reason, the video is presented with interruptions, commonly referred to as freezes. After each freeze, users answer questions about the segment they have just watched and, in some cases, about information from previous segments. This allows us to assess whether the user's SA is progressively built across the different parts of the scenario. The questions, grouped by freeze and SA level, are reported in Table 7.6.

At the start of the scenario, the user's objective is to understand which insulae require pruning of weedy vegetation and which ones have higher priority. The mode is therefore initially set to weedy. During the simulation, however, it starts raining, and the rainfall continues and increases in intensity. One of the key points is to evaluate when, and whether, the user notices this change. In VERGIL, the adaptive interface activates from the first freeze, progressively indicating that it is becoming necessary to check for broken tiles.

In the first segment, the user inspects Area 2 to identify insulae with higher priority for pruning weedy vegetation. From a projection perspective, the user is also expected to determine which insula should be prioritized in case of rain, which in this segment is Insula 15. The user is then asked which area should be addressed next after Area 2, with reference to Area 4, which includes Insula 15 and represents the other key priority after the insulae in Area 2. At each freeze before the last one, the user is asked whether it is time to start checking for broken roofs. This is meant to capture whether the user is beginning to recognize that conditions are changing because of the rain. At this stage, VERGIL provides a low level alert, shown in yellow, suggesting that a change of view may soon be needed. The intent is to prompt readiness without creating undue urgency, since there is not yet an immediate need to switch focus. Users are also asked, at the perception level, whether they observed rain or whether the forecast indicates imminent rainfall.

In the second segment, the user inspects Area 3 to determine whether it requires maintenance with more or less urgency than Area 4, and to identify the insula needing the most urgent attention. In this phase, Area 4 is clearly more vulnerable overall, while within Area 3, Insula 16 is the most at risk. The user is asked whether they can anticipate that Area 4 has the highest vulnerability in terms of weedy vegetation, with Insula 15 as the most critical insula. Meanwhile, the rain related emergency is escalating, and in VERGIL the alert level increases to orange. At this stage, it becomes advisable to start checking for broken tiles to prevent more serious issues. Additional questions address wind conditions and forecasts.

In the third segment, the user inspects Area 4 and identifies which insula requires the most urgent attention. Another question assesses whether the user still recalls which area was most vulnerable within Area 3, which was shown

Freeze	SA level	Question type	Question
	Perception		
		Multiple-choice	Did it rain during the video?
		Multiple-choice	Do the forecasts say it will rain in the next few hours?
F1	**Comprehension**		
		Multiple-choice	Which insula should be given absolute priority for invasive vegetation maintenance, given that it was in the worst condition?
	Projection		
		Multiple-choice	After Area 2 (just shown), which area next requires attention regarding invasive vegetation?
		Multiple-choice	Based on the video, do you think it is necessary to start checking broken roof tiles on the site as well?
	Perception		
		Multiple-choice	Was it windy during the video?
		Multiple-choice	Do the forecasts say it will be windy in the next few hours?
F2	**Comprehension**		
		Multiple-choice	Overall, does Area 3 require more or less urgent invasive-vegetation intervention than Area 4?
		Multiple-choice	In Area 3, which insula requires the most urgent intervention?
	Projection		
		Multiple-response	Even without seeing it in detail yet, which insulae in Area 4 might have priority for invasive vegetation control?
		Multiple-choice	Based on the video, do you think it is necessary to start checking broken roof tiles on the site as well?
	Comprehension		
		Multiple-choice	Among all areas seen so far, which area should be maintained first regarding invasive vegetation?
F3		Open-ended	Write the insula(e) in that area to prioritize for invasive vegetation removal.
		Multiple-response	In Area 3, which insulae require the most urgent intervention?
		Multiple-response	In Area 4, which insulae might have priority for invasive vegetation control?
	Projection		
		Multiple-choice	Based on the video, do you think it is necessary to start checking broken roof tiles on the site as well?
	Perception		
		Multiple-choice	Did it rain during the video?
	Comprehension		
		Multiple-choice	Do the insulae shown in detail require the immediate replacement of roof tiles?
F4		Multiple-choice	Given what you saw, is it feasible to carry out combined interventions on both invasive vegetation and broken roof tiles?
		Open-ended	If yes, on which insulae can these combined interventions be carried out?
		Multiple-response	Which areas still need a closer check for broken roof tiles and damaged waterproof membranes?
	Projection		
		Multiple-choice	Which insula do you expect to be most damaged by the rain?

Table 7.6: Scenario 5 Questionnaire.

earlier. This is intended to evaluate whether VERGIL helps users retain key information longer than the Baseline system. The user is also asked which of all the areas seen so far is most vulnerable in terms of weedy vegetation. Overall, Area 2 remains the highest priority, and within it Insula 15 is the most critical. Given the ongoing worsening conditions, it is again advisable to start checking for broken tiles. Questions about wind and forecasts are also included. In this part, the rainfall becomes heavier and the risk of serious damage increases, so a change of goal is required. In VERGIL, the alert becomes critical and turns red, accompanied by a popup. No perception level questions are included in this segment.

In the final segment, the user's goal changes to identifying maintenance issues related to broken tiles, and the mode is switched accordingly. The user is asked to identify the insulae that should be prioritized from a rainfall perspective. In this case, Insula 15 has the highest priority, followed by Insulae 10 and 8 with lower priority. These latter insulae are also suitable candidates for combined interventions, since they are vulnerable from both perspectives, namely weedy vegetation and broken tiles. Users are also asked whether the tile damage requires immediate replacement, which is expected given the clearly visible breakages, especially in Insulae 15 and 8. They are then asked which insula is likely to suffer the most damage, with Insula 15 as the expected answer given the number of roof sections affected by broken tiles. Finally, users are asked whether it was raining in the video.

Overall, this scenario plays a key role in the evaluation, since it allows us to assess whether VERGIL's adaptive interface and alerting mechanisms provide measurable advantages over the Baseline.

7.4.3 Statistical Significance Evaluation Procedure

We consider multiple usage scenarios indexed by $s \in \{1, \ldots, S\}$, with $S = 5$. For each scenario, each user tested exactly one of two systems/conditions $c \in \{\text{Baseline}, \text{VERGIL}\}$. Therefore, *within each fixed scenario* s, the comparison VERGIL vs. Baseline is a *between-subjects* comparison (independent groups), because each observation belongs to a single condition in that scenario. Across scenarios, however, the same user may appear multiple times (repeated measures across scenarios), and may be assigned to different conditions across different scenarios. This motivates modeling the user effect when performing a pooled (global) analysis across scenarios [275].

7.4.3.1 Dependent Variables

Starting from questionnaire responses / elemental metrics (items), aggregated scores were defined for each user i and scenario s. Let x_{isj} denote the score of user i in scenario s on item j ($j = 1, \ldots, J$). The *global* score for user i and scenario s is defined as:

$$\text{Global-SA}_{is} = \sum_{j=1}^{J} x_{isj}. \tag{7.30}$$

For each level $L \in \{\text{Perception}, \text{Comprehension}, \text{Projection}\}$, letting $\mathcal{J}_L$ be the set of items associated with that level, the level-specific total is:

$$\text{Tot}_{L,is} = \sum_{j \in \mathcal{J}_L} x_{isj}. \tag{7.31}$$

This yields four primary dependent variables (DVs): Global-SA, $\text{Tot}_{\text{Perception}}$, $\text{Tot}_{\text{Comprehension}}$, and $\text{Tot}_{\text{Projection}}$.

7.4.3.2 Descriptive Statistics

For each scenario s and condition c, define the set of users observed in that group:

$$\mathcal{I}_{sc} = \{\, i : \text{ user } i \text{ performed scenario } s \text{ under condition } c \,\}, \qquad n_{sc} = |\mathcal{I}_{sc}|.$$

Let y_{is} be a generic DV observation for user i in scenario s (e.g., $y_{is} = \text{Global-SA}_{is}$). The sample mean and sample standard deviation within group (s, c) are:

$$\bar{y}_{sc} = \frac{1}{n_{sc}} \sum_{i \in \mathcal{I}_{sc}} y_{is}, \tag{7.32}$$

$$s_{sc} = \sqrt{\frac{1}{n_{sc} - 1} \sum_{i \in \mathcal{I}_{sc}} (y_{is} - \bar{y}_{sc})^2}. \tag{7.33}$$

7.4.3.3 Hypothesis Testing

For each DV, the main goal is to test whether, on average, VERGIL improves performance relative to Baseline:

$$H_0 : \mu_{\text{VERGIL}} = \mu_{\text{Baseline}} \qquad \text{vs} \qquad H_1 : \mu_{\text{VERGIL}} \neq \mu_{\text{Baseline}}. \tag{7.34}$$

Since scenarios may differ in inherent difficulty, Scenario is included as a blocking factor in the pooled dataset. Moreover, since the same user appears

across multiple scenarios, a user term is included to control for stable between-user differences.

Let $c(i, s) \in \{\text{Baseline}, \text{VERGIL}\}$ denote the condition assigned to the observation (i, s). The fixed-effects linear model used for each DV is:

$$y_{is} = \mu + \alpha_{c(i,s)} + \beta_s + \delta_i + (\alpha\beta)_{c(i,s),s} + \varepsilon_{is}, \tag{7.35}$$

where μ is the intercept (grand mean), α_c is the main effect of Condition (c), β_s is the main effect of Scenario, δ_i is the subject effect, $(\alpha\beta)_{c,s}$ is the Condition$\times$Scenario interaction, and ε_{is} is an error term. In this model, the main effect of Condition represents the average VERGIL–Baseline difference across scenarios, while the interaction tests whether the VERGIL advantage depends on the scenario.

7.4.3.4 OLS ANOVA

The model is fitted by ordinary least squares (OLS) [276]. Let N be the number of observations used for that DV, and let $\bar{y}$ be the grand mean over all N observations. The *total sum of squares* is:

$$SS_{\text{Tot}} = \sum_{(i,s)} (y_{is} - \bar{y})^2. \tag{7.36}$$

Let $\widehat{y}_{is}$ be the fitted value under the *full* model, and $e_{is} = y_{is} - \widehat{y}_{is}$ the residual. The *residual (error) sum of squares* for the full model is:

$$SS_{\text{Err}} = \sum_{(i,s)} e_{is}^2, \tag{7.37}$$

and the *model* sum of squares is:

$$SS_{\text{Model}} = SS_{\text{Tot}} - SS_{\text{Err}}. \tag{7.38}$$

In our multifactor linear model, each term (Condition, Scenario, Condition$\times$Scenario, and User) is tested with an ANOVA F*-test*, i.e., $F_T \sim F_{df_T, df_{\text{Err}}}$ [276] [277]. To test a given effect T (e.g., Condition, Scenario, Interaction, User), ANOVA computes a sum of squares associated with that term, SS_T. In such a contex, sums of squares are used: intuitively, SS_T quantifies the additional reduction in residual variability attributable to term T when T is added to a model that already contains the other relevant terms. The corresponding degrees of freedom are df_T, while the residual degrees of freedom for the full model are:

$$df_{\text{Err}} = N - p,$$

where p is the number of estimated parameters in the full model.

Mean squares are defined as:

$$MS_T = \frac{SS_T}{df_T}, \qquad MS_{\text{Err}} = \frac{SS_{\text{Err}}}{df_{\text{Err}}}. \tag{7.39}$$

The F-test [276, 277] for testing term T is:

$$F_T = \frac{MS_T}{MS_{\text{Err}}} = \frac{SS_T/df_T}{SS_{\text{Err}}/df_{\text{Err}}}, \tag{7.40}$$

Accordingly, the reported p-value is computed from the F distribution $F_{df_T, df_{\text{Err}}}$:

$$p = P(F_{df_T, df_{\text{Err}}} \geq F_{T,\text{obs}}). \tag{7.41}$$

Considering $T = Condition$ (the VERGIL/Baseline system effect), using a significance level $\alpha = 0.05$, we reject H_0 for the Condition effect if $p < 0.05$, concluding that VERGIL and Baseline differ significantly on that DV in the pooled analysis.

7.4.3.5 Scenario-specific Comparisons

It is also natural to test VERGIL vs. Baseline separately within each scenario s. In that case, the model reduces to a between-subjects comparison within the scenario:

$$y_{is} = \mu_s + \alpha_{c(i,s)} + \varepsilon_{is}, \tag{7.42}$$

and ANOVA assesses whether the two conditions differ significantly in that scenario. With two conditions, the one-way ANOVA within a scenario is equivalent to an independent-samples t-test, since $F_{1,df} = t^2_{df}$; in such a way the p-value can be calculated as equation 7.41 [277, 278].

7.4.3.6 Effect Sizes

Beyond statistical significance, effect sizes were computed. For ANOVA, an effect-size measure for the Condition term can be defined via eta-squared. In multifactor designs, a common choice is *partial eta-squared*:

$$\eta_p^2 = \frac{SS_{\text{Cond}}}{SS_{\text{Cond}} + SS_{\text{Err}}}, \tag{7.43}$$

which represents the proportion of *residual-plus-effect* variance attributable to Condition, within the fitted model.

For scenario-specific (independent-groups) comparisons, Cohen's d [279] was computed as:

$$d = \frac{\bar{y}_{\text{VERGIL}} - \bar{y}_{\text{Baseline}}}{s_p},$$
$$s_p = \sqrt{\frac{(n_{\text{VERGIL}} - 1)s^2_{\text{VERGIL}} + (n_{\text{Baseline}} - 1)s^2_{\text{Baseline}}}{n_{\text{VERGIL}} + n_{\text{Baseline}} - 2}}. \tag{7.44}$$

The sign indicates direction (e.g. $d > 0$ implies higher scores under VERGIL if higher is better), while the magnitude indicates the strength of separation between the two group distributions.

7.4.4 Results and Discussion

At the end of the experimental campaign, we collected the results relating to the scores achieved by users and analysed them from various points of view. As we know, the main metrics that were calculated in relation to the score relate to the averages of the scores achieved on the Baseline and VERGIL systems and the relative standard deviation. This mean and standard deviation were calculated separately on the total scores and on the questions relating to the three levels.

7.4.4.1 Global Results

Analyzing the metrics related to global SA, we note a significant improvement across all scenarios, as can be seen from the barplot in Figure 7.22

Overall, VERGIL achieves higher *Global SA* scores in all scenarios, with particularly large absolute gains in Scenario 5 (Baseline: 12.75 ± 2.26, VERGIL: 16.49 ± 2.07) and clear improvements also in Scenarios 1, 2, and 4, while Scenario 3 shows a more moderate increase.

In terms of *Perception*, as can be seen from Figure 7.23, differences are small and often negligible, with identical means in Scenarios 1 and 2, only minor changes in Scenarios 3 and 5, and a slight decrease in Scenario 4, which is consistent with the overall limited effect observed on this component. This result was expected because, as noted above, our methodological approach primarily supports the understanding and projection phases by introducing a situation model based on CHVIs and the oscillatory goal suggestion model.

By contrast, *Comprehension*, as can be seen from Figure 7.24, improves consistently under VERGIL across all scenarios, with especially pronounced gains in Scenario 3 and a large absolute increase in Scenario 5 (Baseline: 5.94 ± 1.35, VERGIL: 8.15 ± 1.26).

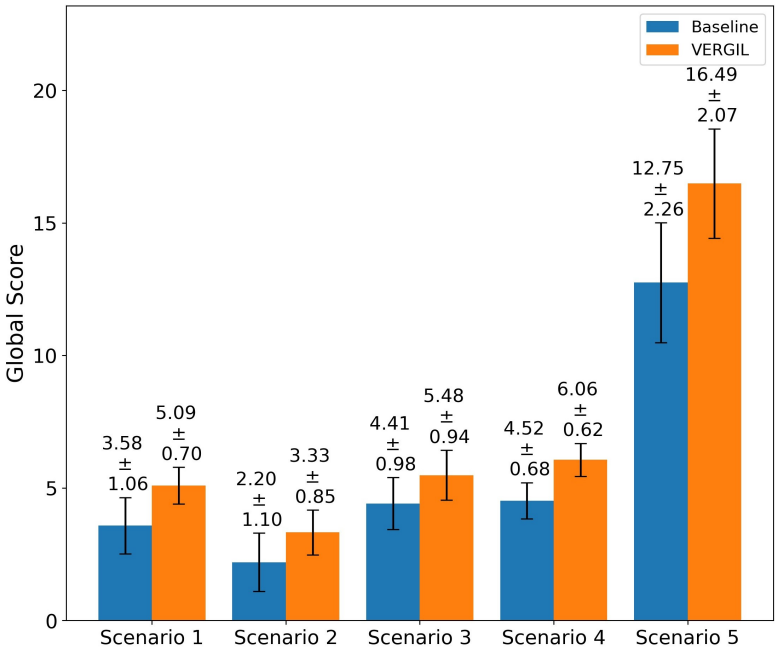


Figure 7.22: Global SA Per Scenario Barplot

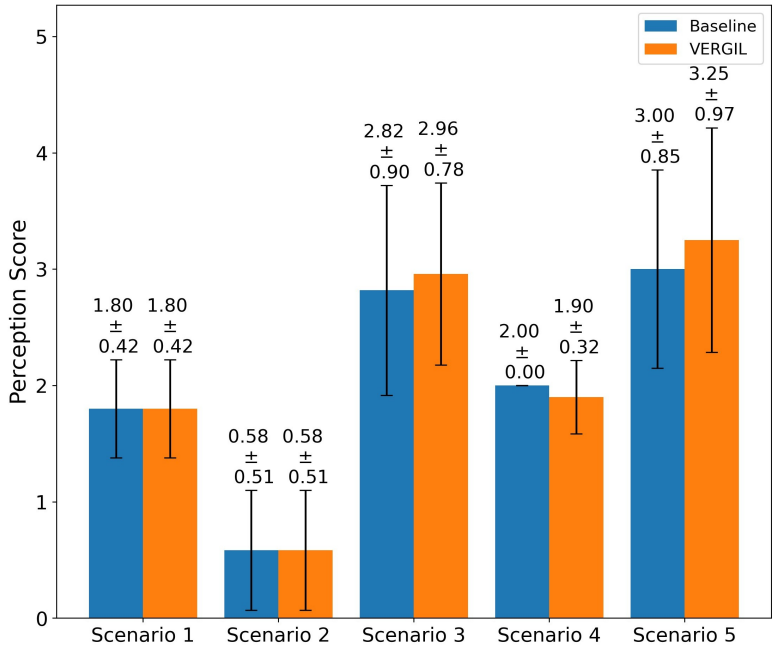


Figure 7.23: Perception SA Per Scenario Barplot

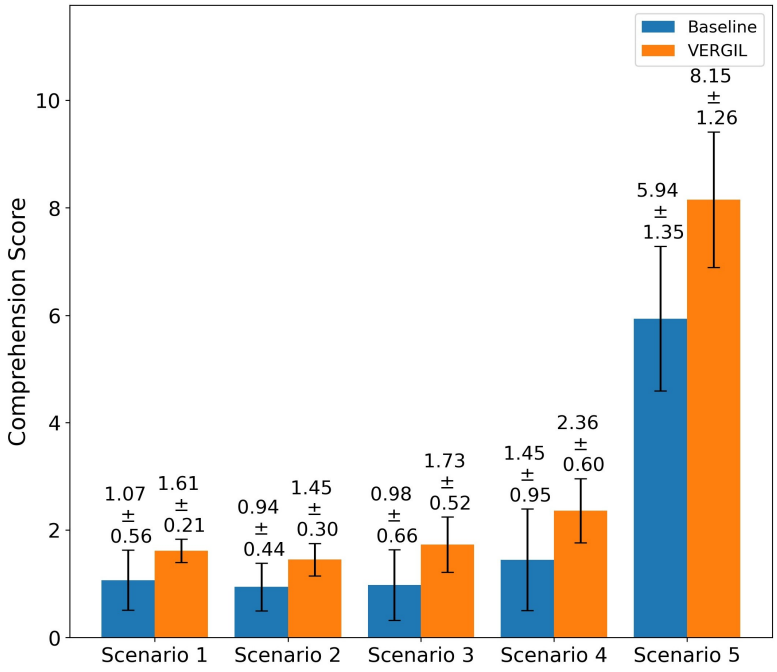


Figure 7.24: Comprehension SA Per Scenario Barplot

A similar pattern emerges for *Projection* (Figure 7.25), where VERGIL outperforms Baseline in every scenario, with strong gains in Scenarios 1 and 2 and a substantial increase in Scenario 5 (Baseline: 3.81 ± 1.01, VERGIL: 5.08 ± 0.81).

Overall, the mean *Global SA* improvement ranges from approximately $+24.3\%$ to $+51.4\%$ across scenarios (average $+32.7\%$), while *Comprehension* improves by roughly $+37.2\%$ to $+76.5\%$ (average $+47.4\%$) and *Projection* by about $+29.0\%$ to $+136.6\%$ (average $+54.8\%$), highlighting that the main benefits of VERGIL emerge at the *Comprehension* and *Projection* levels rather than in *Perception*, but in any case, for *Global*, *Comprehension* and *Projection* the extent of the improvement varies greatly from scenario to scenario, even if it is always present.

Table 7.7: Global OLS-ANOVA p-values

DV	p_{System}	p_{Scenario}	$p_{\text{System}\times\text{Scenario}}$	p_{ID}
Global	5.60×10^{-12}	3.58×10^{-51}	7.11×10^{-3}	5.73×10^{-3}
Perception	7.25×10^{-1}	1.06×10^{-26}	9.65×10^{-1}	3.58×10^{-4}
Comprehension	3.68×10^{-9}	2.37×10^{-44}	2.27×10^{-3}	1.49×10^{-1}
Projection	9.80×10^{-9}	8.07×10^{-35}	1.34×10^{-1}	6.83×10^{-1}

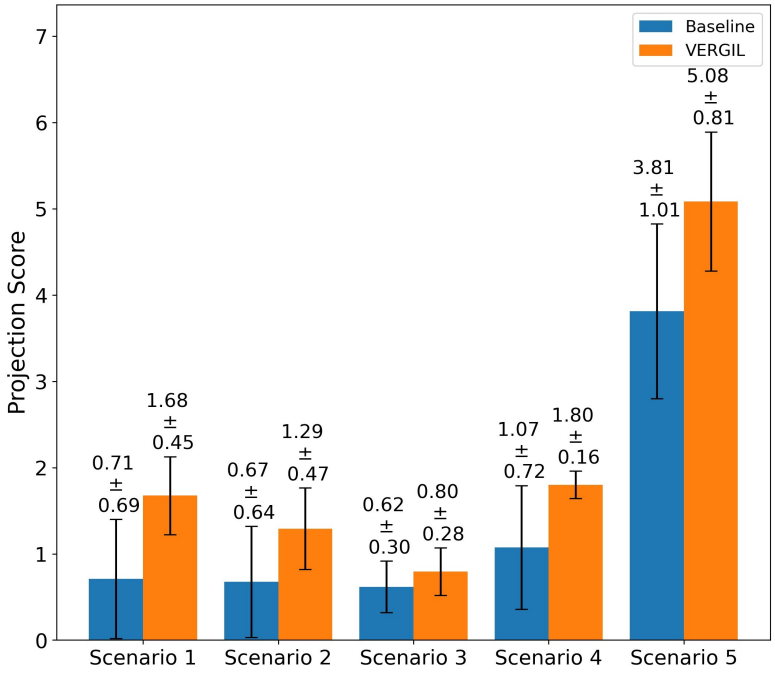


Figure 7.25: Projection SA Per Scenario Barplot

As mentioned above, we performed fixed effects tests using an OLS ANOVA to assess the significance of the results. From the Table 7.7, we can interpret the outcomes through the p values associated with each term in Equation 7.35.

The term p_{System} captures the statistical significance of the effect of the system used (Baseline vs VERGIL). When it is below 0.05, and especially when it is very small, we can reject the null hypothesis that the two systems yield the same average scores on *Global SA* and on the *Comprehension* and *Projection* components. In our results, p_{System} is very small for *Global SA*, *Comprehension*, and *Projection*, indicating statistical significance. This is not the case for *Perception*. Here, the mean scores differ only marginally, so we expected the effect not to be statistically significant.

Beyond significance, for the *Global SA* DV, the classical eta squared for the System effect was $\eta^2 = 0.0358$. This indicates that, within the fitted model, which controls for Scenario, User, and their interaction, a non trivial portion of the residual plus effect variance is attributable to the *System Used* factor. Consistently, the equivalent Cohen effect size was large ($d_{\text{eq}} = 1.838$), suggesting a strong separation between VERGIL and Baseline performance on overall SA.

For the *Perception* DV, the effect sizes were negligible ($\eta^2 = 0.000268$, $\eta_p^2 = 0.001601$, $d_{\text{eq}} = 0.080$), in line with the non significant System p

value. This indicates that VERGIL does not provide a practically meaningful improvement on the Perception component in the pooled analysis.

For *Comprehension*, the System effect had a non trivial magnitude, with $\eta^2 = 0.0359$ and a large Cohen effect size ($d_{\text{eq}} = 1.505$). This suggests that VERGIL provides a substantial overall improvement in Comprehension, although the significant System$\times$Scenario interaction indicates that the size of this benefit varies across scenarios and should be examined through scenario specific contrasts.

Finally, for the *Projection* DV, effect sizes were also sizeable ($\eta^2 = 0.0580$, $d_{\text{eq}} = 1.454$), indicating a large overall improvement of VERGIL on Projection. Unlike Comprehension, the non significant interaction suggests that this improvement is comparatively more stable across scenarios in the pooled analysis.

Looking at $p_{Scenario}$, the very small values indicate that the results obtained by the users varies substantially from scenario to scenario, both overall and at each SA level. This is expected, since the number of questions and the difficulty of the scenarios differ considerably.

For the interaction term $p_{\text{System}\times\text{Scenario}}$, the low values for Perception and Comprehension indicate that the improvements of VERGIL vary noticeably across scenarios. For this reason, it is advisable to evaluate results and statistical significance on a per scenario basis, in order to better understand the improvements achieved in each case.

Finally, the low p_{User} for *Global SA* and *Perception* indicates substantial variability across participants. This variability is largely driven by Perception, meaning that some users are much better than others at noticing the cues targeted by perception questions. By contrast, this variability does not appear to meaningfully affect Comprehension and Projection, which is reassuring. It suggests that the system supports users more uniformly at these levels, something we further verify through a dedicated user level analysis.

7.4.4.2 In-depth Scenario Analysis

As written before, the significant $System \times Scenario$ interaction implies that the magnitude of this benefit varies across scenarios and should therefore be investigated through a scenario-specific analysis, which we do below.

In Scenario 1, the barplot in Figure 7.26 show that the *Total* score of VERGIL increases when compared to Baseline from 3.58 ± 1.06 to 5.09 ± 0.70, a gain of $+1.51$ points (i.e., $+42.2\%$). Similarly, *Comprehension* rises from 1.07 ± 0.56 to 1.61 ± 0.21, a gain of $+0.54$ points (i.e., $+50.5\%$), and *Projection* improves from 0.71 ± 0.69 to 1.68 ± 0.45, a gain of $+0.97$ points (i.e., $+136.6\%$). In this case, the improvements are attributable to VERGIL users' improved ability to

recognize the areas and islands with the highest priority for weedy vegetation removal, as well as to understand which areas to focus on next.

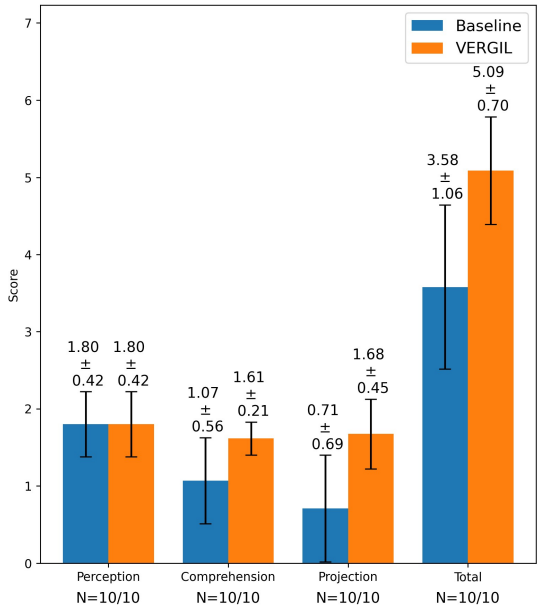


Figure 7.26: Scenario 1 Score Barplot by Level

In Scenario 2, the barplot in Figure 7.27 show that the *Total* score of VERGIL increases from 2.20 ± 1.10 to 3.32 ± 0.85, a gain of $+1.12$ points (i.e., $+50.9\%$). Likewise, *Comprehension* increases from 0.94 ± 0.44 to 1.45 ± 0.30, a gain of $+0.51$ points (i.e., $+54.3\%$), and *Projection* rises from 0.67 ± 0.64 to 1.29 ± 0.47, a gain of $+0.62$ points (i.e., $+92.5\%$). In this case, there were no major improvements in identifying the areas with the highest density of damaged roofs; however, VERGIL users were able to pinpoint the specific insulae where replacement was most urgent. They were also better able to recognize the importance of maintaining insula 15 due to broken tiles, even though this was not their current goal.

In Scenario 3, the barplot in Figure 7.28 show that the improvements of VERGIL are more moderate for *Total* and *Projection*, with *Total* moving from 4.41 ± 0.98 to 5.48 ± 0.94, a gain of $+1.07$ points (i.e., $+24.3\%$), and *Projection* increasing from 0.62 ± 0.30 to 0.80 ± 0.28, a gain of $+0.18$ points (i.e., $+29.0\%$). By contrast, *Comprehension* shows a particularly large relative increase, from 0.98 ± 0.66 to 1.73 ± 0.52, a gain of $+0.75$ points (i.e., $+76.5\%$).

Here, baseline users are often able to identify at least one, and often two, of the insulae where conduits replacement is most urgent; however, they find it difficult to understand that the insula requiring replacement most urgently, due to environmental factors and the greater number of conduits, is 9. From this

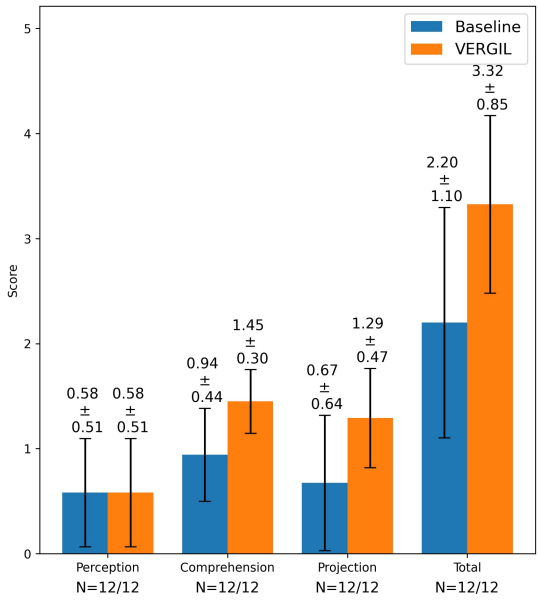


Figure 7.27: Scenario 2 Score Barplot by Level

perspective, VERGIL improves users' performance. Furthermore, VERGIL users also show an improvement in understanding how insula 15, and broken tiles in general, could pose a problem in the event of rain, as much as, and even more than, conduits, even though this was not the main goal.

In Scenario 4, from the barplot in Figure 7.29, VERGIL again shows clear advantages, with the *Total* score increasing from 4.52 ± 0.68 to 6.06 ± 0.62, a gain of $+1.54$ points (i.e., $+34.1\%$), *Comprehension* rising from 1.45 ± 0.95 to 2.36 ± 0.60, a gain of $+0.91$ points (i.e., $+62.8\%$), and *Projection* increasing from 1.07 ± 0.72 to 1.80 ± 0.16, a gain of $+0.73$ points (i.e., $+68.2\%$). In this scenario, improvements are observed in VERGIL users' ability to identify the insulae with the highest priority for the removal of weedy vegetation (Insula 10, followed by 7), as well as the imminent emergency of safeguarding zone 4 and, in particular, insula 15, which is the most vulnerable due to broken tiles. There are also improvements in identifying areas where combined interventions can be carried out on issues such as weedy vegetation and broken tiles.

Finally, Scenario 5, from the barplot in Figure 7.30, VERGIL exhibits the largest absolute gains with respect to Baseline, with VERGIL *Total* score increasing from 12.75 ± 2.26 to 16.49 ± 2.07, a gain of $+3.74$ points (i.e., $+29.3\%$), *Comprehension* rising from 5.94 ± 1.35 to 8.15 ± 1.26, a gain of $+2.21$ points (i.e., $+37.2\%$), and *Projection* increasing from 3.81 ± 1.01 to 5.08 ± 0.81, a gain of $+1.27$ points (i.e., $+33.3\%$).

With regard to this scenario, the various freezes indicate that VERGIL users

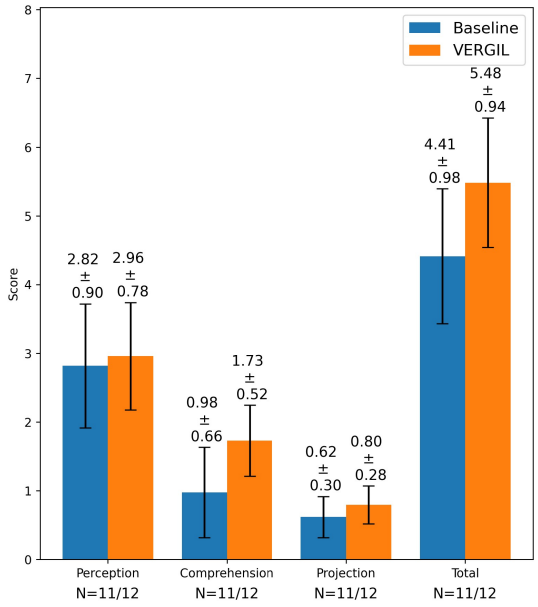


Figure 7.28: Scenario 3 Score Barplot by Level

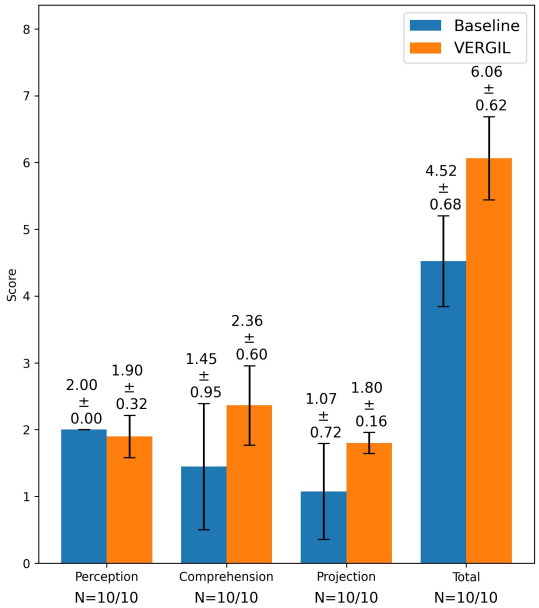


Figure 7.29: Scenario 4 Score Barplot by Level

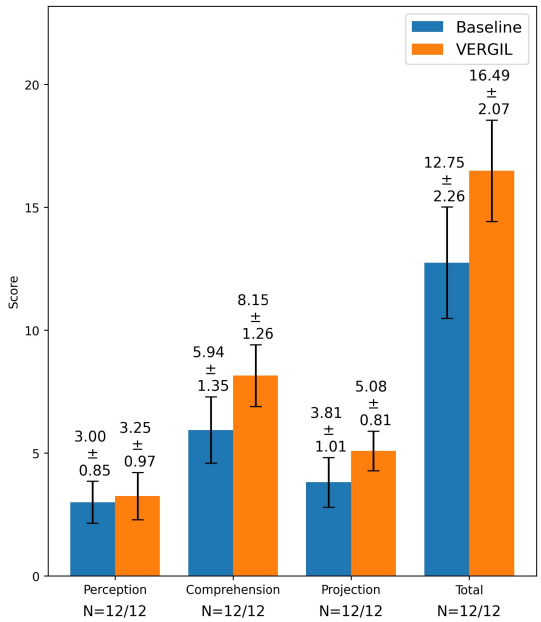


Figure 7.30: Scenario 5 Score Barplot by Level

improve their ability to understand which areas are most sensitive in terms of weedy vegetation removal, as well as which individual insulae are most critical. During the simulation, a rain emergency arises, and VERGIL users are able to detect it earlier and to understand, slightly in advance, when it becomes necessary to check for broken tiles, without this being so invasive as to distract them. Accordingly, VERGIL alert functionalities allow users to improve their performance in determining the most vulnerable and at-risk areas with respect to both the initial and final goals, namely weedy vegetation and broken tiles, and they are also better able to identify areas where interventions for both maintenance issues can be carried out in a combined manner.

The color-based vulnerability interface related to Goal-driven CHVIs not only helps users notice otherwise complicated details in areas other than those they are focusing on, but also allows them to remember areas with a certain vulnerability over time, between freezes, so that SA increases and they can obtain a quicker and more effective overview of the entire situation. Between freezes, VERGIL users can more effectively compare the vulnerabilities of the area they are looking at with those observed previously.

In Table 7.8, the results in terms of statistical significance and effect size are reported, with the bold p-values highlighting the scenarios and dependent variables that pass the significance test.

Overall, the improvement of VERGIL in terms of *Total* score is statistically

Scenario	dv	p	η^2	d
1	Total	$\mathbf{1.40 \times 10^{-3}}$	4.39×10^{-1}	1.68×10^{0}
2	Total	$\mathbf{1.01 \times 10^{-2}}$	2.64×10^{-1}	1.15×10^{0}
3	Total	$\mathbf{1.45 \times 10^{-2}}$	2.53×10^{-1}	1.11×10^{0}
4	Total	$\mathbf{5.20 \times 10^{-5}}$	6.07×10^{-1}	2.36×10^{0}
5	Total	$\mathbf{3.49 \times 10^{-4}}$	4.48×10^{-1}	1.73×10^{0}
1	Perception	1.00×10^{0}	4.33×10^{-32}	0×10^{0}
2	Perception	1.00×10^{0}	1.98×10^{-32}	0×10^{0}
3	Perception	6.94×10^{-1}	8.00×10^{-3}	1.67×10^{-1}
4	Perception	3.31×10^{-1}	5.30×10^{-2}	-4.47×10^{-1}
5	Perception	5.08×10^{-1}	2.00×10^{-2}	2.74×10^{-1}
1	Comprehension	$\mathbf{9.90 \times 10^{-3}}$	3.16×10^{-1}	1.29×10^{0}
2	Comprehension	$\mathbf{3.50 \times 10^{-3}}$	3.27×10^{-1}	1.34×10^{0}
3	Comprehension	$\mathbf{5.80 \times 10^{-3}}$	3.09×10^{-1}	1.28×10^{0}
4	Comprehension	$\mathbf{1.85 \times 10^{-2}}$	2.71×10^{-1}	1.16×10^{0}
5	Comprehension	$\mathbf{4.16 \times 10^{-4}}$	4.39×10^{-1}	1.70×10^{0}
1	Projection	$\mathbf{1.60 \times 10^{-3}}$	4.32×10^{-1}	1.65×10^{0}
2	Projection	$\mathbf{1.39 \times 10^{-2}}$	2.45×10^{-1}	1.09×10^{0}
3	Projection	1.53×10^{-1}	9.50×10^{-2}	6.19×10^{-1}
4	Projection	$\mathbf{5.90 \times 10^{-3}}$	3.51×10^{-1}	1.40×10^{0}
5	Projection	$\mathbf{2.60 \times 10^{-3}}$	3.45×10^{-1}	1.39×10^{0}

Table 7.8: Per-scenario ANOVA OLS VERGIL vs Baseline

significant in all scenarios ($p < 0.05$), with consistently very large effects (Cohen's $d \geq 1.11$) and substantial η^2 values. Scenario 4 stands out in particular ($\eta^2 = 0.607$, $d = 2.358$), indicating an extremely strong separation between the VERGIL system score and the baseline score.

Comprehension is also significant in every scenario (all $p < 0.05$), with large and consistent effect sizes (d between 1.16 and 1.70, and η^2 between 0.271 and 0.439), suggesting a robust and stable benefit.

For *Projection*, differences are significant in Scenarios 1, 2, 4, and 5 (all $p < 0.05$), again with large effects (d between 1.09 and 1.65). In Scenario 3, instead, the effect is not significant ($p = 0.153$) and the magnitude is more limited (moderate, $d = 0.619$, $\eta^2 = 0.095$), suggesting that *Projection* gains are more scenario dependent.

Finally, *Perception* shows no significant differences in any scenario (all $p > 0.05$), and effect sizes are negligible or small, consistent with the absence of a practically meaningful improvement on this component.

7.5 Summary

In this chapter we have presented the final platform, named VERGIL, that integrates the components introduced in the previous chapters into a single operator oriented system for archaeological site protection, developed and tested in the Pompeii Archaeological Site.

The chapter starts from a goal directed task analysis carried out with domain experts to identify operators goals, the decisions they must take, and the information they need to take those decisions. The resulting hierarchy frames the overall mission of safeguarding the site and highlights the need to balance protection actions with site usability. The design is grounded on the branch related to ensuring a quick and effective response to danger, which is operationalized through three main activities: assessing maintenance issues on structures, assessing risks driven by contextual factors such as weather and visitor presence, and establishing priorities to update the maintenance plan. For each activity the analysis clarifies which information must be available at the levels of perception, comprehension, and projection, and this directly drives the interface requirements and the way information is presented.

On this basis, the chapter describes the implementation of the platform and its interface, centered on an interactive georeferenced map of the monitored area and a set of widgets that expose contextual variables. The map overlays the site orthophoto with area level vulnerability information computed in real-time by fusing detected maintenance issues with contextual signals such as rainfall, wind, temperature, humidity, and visitor levels. Vulnerability is shown through

an intuitive color coding of the site sub-areas and is complemented by labels that provide key attributes needed to interpret the map and plan interventions. To reduce reliance on memory and to support informed decisions when inspecting different parts of the site, the platform includes a structured knowledge base derived from site data and expert input, which can be queried from the interface to retrieve information about areas and structures and to contextualize what is observed on the map.

A distinctive aspect of the interface is that it is goal-driven and adaptive. Different visualization modes allow operators to focus on specific classes of maintenance issues, and the vulnerability computation is adjusted accordingly so that it remains aligned with the current task. At the same time, the system addresses the risk that operators become overly focused on a single goal by introducing progressive alerts that guide attention when contextual conditions indicate an increasing risk outside the current focus. In particular, when rainfall intensifies or persists while the operator is inspecting other issues, the interface gradually increases the alert level and prompts a shift toward checking roof damage, using an escalation strategy that aims to be noticeable without being unnecessarily disruptive.

The chapter concludes with an evaluation based on SAGAT aimed at assessing how well the platform supports operators across perception, comprehension, and projection. The results show a consistent improvement over a baseline interface that provides access to the same underlying data but lacks explicit vulnerability synthesis and adaptive alerting. Improvements are especially evident for comprehension and projection, where the integration of heterogeneous cues, the prioritization support, and the guidance during goal changes are most critical, while the effect on perception is limited. Overall, the findings support the conclusion that representing site conditions through an integrated vulnerability view, combined with an adaptive and knowledge-supported interface, provides measurable benefits for operator decision-making in both routine maintenance planning and evolving emergency situations.

Chapter 8

Conclusions and Future Works

"Whenever I'm about to do something, I think, "Would an idiot do that?" And if they would, I do not do that thing."

— Dwight Schrute, *The Office*

This chapter concludes the thesis with a brief summary and final remarks on the main contributions of the research work, highlighting how these contributions address the objectives set on the basis of the challenges identified (Chapter 1), and how our work supports the SDGs mentioned at the beginning. We also discuss reflections on possible future developments.

8.1 Summary

This thesis builds on the idea that sites in the cultural heritage domain can be treated as Cultural Heritage Critical Infrastructure (CHCI), where maintenance must be timely, planned, and carried out by human operators. Cultural heritage protection is therefore a human-in-the-loop challenge. Automation can support operators in understanding ongoing conditions, but decisions and prioritization must remain under human control, since interventions must be justified and assessed by domain experts, who are also responsible for protecting the property assigned to them. There is therefore also a question of accountability that must not be overlooked.

Because physical, technological, and human dimensions continuously interact, we frame modern CHCIs as Cyber-Physical-Social Systems (CPSSs). This framing enables the formalization of interactions among actors and the defini-

tion of CPSS functionalities through shared standards, supporting replicability across different CHCIs. By modelling CPSSs according to situation awareness principles, we strengthen operators' decision-making capabilities.

We first review predictive maintenance in CHCIs, define a reference maintenance lifecycle, and discuss existing vulnerability indices. We then introduce situation awareness and analyse the factors that degrade it in human–CPSS interactions. Next, we present an overview of the state-of-the-art in AI for cultural heritage, highlighting techniques, data sources, and addressed problems, with particular attention to the detection of maintenance issues in CHCIs.

We propose a novel Situation-aware Cyber-Physical-Social System methodological approach for cultural heritage protection, designed according to the Endsley model. Since several studies have shown that, under the same conditions, a higher SA among human operators leads them to make better decisions on average [20] [21] [22], the approach is designed to assist operators in their decision-making by directly supporting their SA. The Cyber-Layer is structured around the levels of situation awareness so that it can both build situational knowledge and share it with users, effectively supporting their decisions. Within this CPSS, we introduce a goal-driven Cultural Heritage Vulnerability Index (CHVI) to formalize situations within CHCIs. Unlike traditional vulnerability measures based on static and infrequent surveys, CHVI is computed in real-time by combining automatically detected maintenance issues with contextual environmental and social information, providing an up-to-date representation of site conditions. Moreover, vulnerability is computed according to the user's current goal, enabling personalized support for routine activities. A proof of concept demonstrates feasibility through case studies at Pompeii, Paestum, and the Colosseum.

Following the CPSS architecture, we describe the main system components, proposing novel approaches and techniques.

Starting from physical and virtual sensors, data sources, and the preprocessing pipeline, the CPSS supports both a fully automatic workflow, including post-classification data cleaning, and a semi-automatic, expert-driven workflow that processes raw data into virtual objects such as 2D orthophotos and 3D models.

We propose GARDA: a Granular Association-rule-based Data Imputation Approach, a lightweight method for imputing missing data in small physical and virtual sensor networks. GARDA leverages granular computing by discretizing continuous data through granulation, applying granular association rules, and completing the process with a final traditional imputation step.

Two AI solutions for maintenance-issue identification in archaeological sites are presented, namely FAUNO, a YOLO-based detection approach, and HydraML: a Hybrid Damaged Roofs AutoML Classifier for Cultural Heritage. The focus is placed on HydraML, as it achieves the best performance on

roof-related issues, which are particularly critical in archaeological contexts. HydraML combines deep feature extraction with non-deep AutoML models to reduce training time and energy consumption, and it is validated on two Pompeii roof datasets.

Finally, VERGIL is proposed as a SA-oriented platform for archaeological site protection, featuring an adaptive interface with a novel Oscillatory Goal Suggestion approach that supports both routine and emergency maintenance and escalates alerts when urgent attention is required. VERGIL is in fact an instance of the entire SA-CPSS methodology applied to the archaeological site of Pompeii. The whole SA-CPSS methodological approach is evaluated, through VERGIL, using SAGAT to assess users' situation awareness and decision-making performance in dedicated Pompeii-based scenarios.

8.2 Final Remarks

In Chapter 1, we highlighted a set of challenges related to a concrete operational gap, which emerged in particular from the analysis conducted at the archaeological site of Pompeii. These considerations also proved relevant in other contexts, such as the archaeological site of Paestum and of the Colosseum. CHCI maintenance requires continuous monitoring and prioritization, yet fully manual approaches are slow, costly, and error-prone, while fully automated solutions are unrealistic given the complexity and uniqueness of cultural heritage contexts. Moreover, both in cultural heritage and, more broadly, in critical infrastructure settings, accountability constraints make full automation not feasible.

Within this context, we proposed a methodological approach aimed at maximizing the effectiveness of operators' decisions in CHCI maintenance. We therefore outline how our research objectives were achieved and how the associated challenges were addressed:

Objective 1. Analyse state-of-the-art approaches and techniques to support human operators' decision-making in the maintenance of CHCIs: This objective primarily addresses **Challenge 1**, which emphasizes the infeasibility of full automation and the need to keep operators meaningfully involved. The thesis responds by first analyzing how CHCI maintenance is currently carried out, including workflows, regulatory constraints, and common sources of error. By framing cultural heritage sites as Cultural Heritage Critical Infrastructures, the work clarifies why decision support must preserve accountability and human judgment. The analysis of situation awareness demons enables

the identification of the most critical issues in this domain and the development of targeted solutions. This foundational analysis provides the conceptual grounding for all subsequent contributions and ensures that the proposed solutions are aligned with real operational needs rather than abstract automation goals. This objective is achieved through the formalization of the CHCI Maintenance Activity Lifecycle (**Contribution 1**), which serves as the reference model for the rest of the thesis, and through the analysis of AI for cultural heritage protection (**Contribution 2**), which lays the foundation for the development of enabling technologies within the proposed methodological approach.

Objective 2. **Define situations within a CHCI and manage them through a novel methodological approach that supports users' decision-making:** This objective directly addresses **Challenge 1** and **Challenge 2**, namely the need to support operators without removing them from the loop and to enable effective recognition of risk situations. The thesis meets this objective by proposing a Situation-aware Cyber-Physical-Social System methodological approach (**Contribution 3**) and by introducing a goal-driven notion of situation tailored to CHCI maintenance (**Contribution 4**). The CPSS provides a general and replicable methodological approach for supporting operators' situation awareness. It enables rapid perception of environmental elements, as well as their comprehension and projection, which are essential for allowing users to focus on maintenance planning and on understanding intervention priorities within the CHCI. A key aspect of this process is the definition of risk situations, which in the cultural heritage domain is achieved by modelling the vulnerability of different areas of a site through indices that account for multiple factors related to site condition and characteristics. To the best of our knowledge, this work is the first to link vulnerability assessment to the dynamic synthesis of automatically detected maintenance issues with contextual information such as weather conditions and visitor flows, which are crucial for accurately capturing risk across site areas. Incorporating the user's current goal into this process further enhances the usefulness of the approach for operators working in CHCIs.

Objective 3. **Address the challenge of low-quality and missing sensor data by reconstructing missing information and reducing perception-level errors:** This objective is motivated by **Challenge 3**, which identifies unreliable and incomplete sensing as a

critical threat to perception and, consequently, to higher-level reasoning. The thesis addresses this challenge by strengthening the sensing layer before data are consumed by AI or decision-support components. The proposed solution, GARDA, is a lightweight method for reconstructing missing data in small physical and virtual sensor networks (**Contribution 5**). By reducing reconstruction errors under increasing levels of missing data, GARDA acts as a protective layer that stabilizes perception and prevents error propagation to vulnerability assessment and decision-making processes. GARDA improves the performance of traditional imputation techniques while approaching, and in some cases surpassing, deep learning approaches. At the same time, it avoids the high storage and computational requirements of deep learning models, which are often impractical in IoT and edge computing environments that are common in large and complex CHCI deployments.

Objective 4. Automate the identification of maintenance issues in CHCIs using a sustainable, energy-efficient AI approach: This objective addresses **Challenge 4**, which concerns the difficulty of fast issue detection in complex CHCI environments. The thesis meets this objective by introducing AI-based perception modules that support, rather than replace, human inspection activities. FAUNO demonstrates the detection of maintenance issues in complex environments, while HydraML focuses on robust and sustainable classification of roof-related maintenance issues (**Contribution 6**). HydraML combines deep feature extraction with non-deep AutoML models to reduce training time and energy consumption, presenting an effective, efficient, and sustainable solution. This methodological approach is specifically designed to mitigate the environmental impact associated with training deep neural networks, an aspect that can no longer be overlooked or sacrificed solely for economic or performance gains. The relevance of sustainability is further underscored by the United Nations Agenda, in which several Sustainable Development Goals address this issue, including SDG 16. Compared to traditional deep and hybrid solutions, as well as fully deep AutoML approaches, HydraML demonstrates superior performance in terms of both computational efficiency and classification accuracy on real datasets from the archaeological site of Pompeii. Moreover, it can be regarded as a reusable methodological approach that is applicable beyond cultural heritage to

other domains and problem settings.

Objective 5. Ensuring that operators are consistently prepared to respond to emergencies, while maximizing their performance both in routine activities and when facing imminent risks: This objective directly addresses **Challenge 5**, which highlights the risk of attentional tunneling and cognitive overload in data-rich monitoring systems. The thesis fulfills this objective through VERGIL, a SA-oriented platform for archaeological site protection featuring a novel adaptive interface that supports both routine and emergency maintenance and escalates alerts when urgent attention is required (**Contribution 7**). The adaptive interface exploits a novel Oscillatory Goal Suggestion approach that gradually signals to the user when it is time to switch goals as an emergency occurs or becomes imminent. The model is inspired by the brain's alertness states during attentional shifts. Building on this mechanism, the interface can guide users toward a goal change when necessary, for example when a current or impending emergency arises, even while they are engaged in other tasks, while avoiding unnecessary disturbances when intervention is not required. Users always retain control over shifts of attention, with VERGIL providing guidance only when appropriate and without causing undue alarm. The SAGAT-based evaluation demonstrates that this approach improves comprehension and projection in real-world scenarios related to the archaeological site of Pompeii, confirming that the platform supports higher levels of situation awareness without overwhelming users. This objective also contributes to addressing **Challenge 1**, as VERGIL represents a concrete platform implementation that supports operators in their activities while allowing them to remain fully in control of their decisions.

8.3 Limitations

We believe it is useful to discuss some limitations identified in the approaches proposed in this thesis.

Regarding the data imputation approach, the analyses in Section 5.3.2 show that the method is effective for small sensor networks. Since it remains Apriori-like, it is best suited to highly correlated sensors. In low-correlation settings, where rule-based approaches struggle, its performance was only competitive with deep learning, while SAITS significantly outperformed GARDA.

With respect to automatic maintenance-issue identification, HydraML, described in Section 6.3, achieves strong results in terms of both accuracy and computational cost, improving over existing work. A major limitation, however, is that it requires a priori segmentation of structures. While this is reasonable in relatively low-variability contexts such as cultural heritage sites, it still entails a non-negligible effort, especially during the initial setup.

The definition of the CHVI vulnerability index can also be improved. In its current form, it does not account for damage severity, which HydraML can estimate and which could be crucial for correctly prioritizing interventions. To further enhance effectiveness, CHVI should also incorporate more detailed information about areas and structures, including the materials from which they are composed. The definition of contributions is also inherently complex, and data-driven approaches would be beneficial for their elicitation.

Finally, the proposed Oscillatory Goal Suggestion approach in section7.3.3.2 is implemented here using a threshold-based system. This requires additional effort to adapt to each domain and deployment context and often depends on expert input, which may not always be available.

8.4 Future Works

Although the objectives defined at the beginning of the thesis have been achieved, their accomplishment has highlighted several areas in which further improvements are possible.

With respect to the data imputation method, future work should explore granulation strategies that explicitly account for data uncertainty and vagueness, such as fuzzy sets and rough sets. In addition, alternative reconstruction schemes beyond average-based operators could be investigated to further improve the method's performance.

Regarding the sustainable approach to classifying roof-related maintenance issues, future research will focus on increasing robustness to shadows, challenging lighting conditions, and structural variability in materials and architectural features. To improve generality and applicability, it will also be important to rigorously define stopping strategies for training the deep component that maximize the overall performance of the hybrid approach.

Considering the CHVI calculation, risk situations management could be further improved by incorporating not only the presence but also the severity of the identified maintenance issues, together with more detailed information about the materials and structural characteristics of the assets.

The proposed Oscillatory Goal Suggestion approach implemented in the VERGIL adaptive interface can also be enhanced in several ways. The pro-

posed interface already guides users between goal-driven and data-driven information processing, across different alert states, gradually redirecting attention toward new goals in emergency situations without unnecessarily disrupting ongoing tasks. However, the transitions between these states remain a critical aspect. While threshold-based implementations have proven effective, they are inherently imprecise and may lead to abrupt transitions. Moreover, identifying appropriate thresholds is complex and typically requires expert intervention, as both domains and addressed issues may vary significantly. Future work will therefore explore data-driven learning approaches for managing these transitions, leveraging recent advances in LLMs.

In parallel, visual attention-capturing strategies used to suggest goal changes could also be extended to highlight critical contextual information, such as rain and wind conditions, with the aim of improving users' SA already at the perception stage. A deeper awareness of environmental conditions may further support operators in making more informed and effective decisions.

Bibliography

[1] X. Liang, F. Liu, L. Wang, B. Zheng, and Y. Sun, "Internet of cultural things: Current research, challenges and opportunities," *Computers, Materials & Continua*, vol. 74, pp. 469–488, 2022.

[2] M. Casillo, F. Colace, A. Lorusso, D. Santaniello, and C. Valentino, "Integrating physical and virtual experiences in cultural tourism: An adaptive multimodal recommender system," *IEEE Access*, vol. 13, pp. 28 353–28 368, 2025.

[3] H. Choi and S. Kim, "Proposal of smart guide system for cultural heritage of archaeological sites using iot-focusing on hoeamsa temple site in yangju, korea," *Mathematical Biosciences and Engineering*, vol. 20, no. 5, pp. 8745–8765, 2023.

[4] Council of the European Union, "Council directive 2008/114/ec of 8 december 2008 on the identification and designation of european critical infrastructures and the assessment of the need to improve their protection," Official Journal of the European Union, L 345/75, 23 Dec. 2008 (CELEX: 32008L0114), 2008, art. 2(a). [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32008L0114

[5] I. Anglmayer, "European critical infrastructure: Revision of directive 2008/114/ec," European Parliamentary Research Service (EPRS), Implementation Appraisal, PE 662.604, February 2021, 2021. [Online]. Available: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/662604/EPRS_BRI(2021)662604_EN.pdf

[6] U.S. Congress, "Uniting and strengthening america by providing appropriate tools required to intercept and obstruct terrorism (usa patriot act) act of 2001," Public Law 107–56, Oct. 26, 2001, 2001, see Title X, Sec. 1016(e). [Online]. Available: https://www.congress.gov/107/plaws/publ56/PLAW-107publ56.pdf

[7] U.S. House of Representatives, Office of the Law Revision Counsel, "42 u.s.c. sec. 5195c: Critical infrastructure protection," United States Code, 2025, see subsection (e).

[8] G. D'Aniello, M. Gaeta, and M. Z. Reformat, "Collective perception in smart tourism destinations with rough sets," in *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, 2017, pp. 1–6.

[9] A. Basu, S. Paul, S. Ghosh, S. Das, B. Chanda, C. Bhagvati, and V. Snasel, "Digital restoration of cultural heritage with data-driven computing: A survey," *IEEE Access*, vol. 11, pp. 53 939–53 977, 2023.

[10] A. De Benedictis, F. Flammini, N. Mazzocca, A. Somma, and F. Vitale, "Digital twins for anomaly detection in the industrial internet of things: Conceptual architecture and proof-of-concept," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 12, pp. 11 553–11 563, 2023.

[11] F. Corradini, C. Grigioni, A. Antonucci, J. Guzzi, and F. Flammini, "Experimental evaluation of road-crossing decisions by autonomous wheelchairs against environmental factors," in *Intelligent Transport Systems*, A. Kocian, P. Milazzo, A. L. Henriques Martins, M. Nanni, and L. Pappalardo, Eds. Cham: Springer Nature Switzerland, 2025, pp. 363–380.

[12] S. Proia, R. Carli, G. Cavone, and M. Dotoli, "Control techniques for safe, ergonomic, and efficient human-robot collaboration in the digital industry: A survey," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 1798–1819, 2022.

[13] S. Proia, G. Cavone, P. Scarabaggio, R. Carli, and M. Dotoli, "Safety compliant, ergonomic and time-optimal trajectory planning for collaborative robotics," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 594–605, 2025.

[14] A. Roumana, A. Georgopoulos, and A. Koutsoudis, "Developing an educational cultural heritage 3d puzzle in a virtual reality environment," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B2-2022, pp. 885–891, 2022.

[15] O. Menaguale, *Digital Twin and Cultural Heritage – The Future of Society Built on History and Art*. Cham: Springer International Publishing, 2023, pp. 1081–1111.

[16] M. Maksimović and M. Ćosović, "Preservation of cultural heritage sites using iot," in *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 2019, pp. 1–4.

[17] X. Yang, Z. Fan, H. Ge, T. Michikata, R. Shibasaki, and N. Koshizuka, "Causality-aware next location prediction framework based on human mobility stratification," in *2024 IEEE Smart World Congress (SWC)*, 2024, pp. 1135–1142.

[18] T. M. H. Gedara, V. Loia, and S. Tomasiello, "A fuzzy-based multimodal approach for interpretable fake news detection," *Applied Soft Computing*, vol. 179, p. 113277, 2025.

[19] S. Masini, S. Bacci, F. Cipollini, and B. Bertaccini, "Revealing the structural behaviour of brunelleschi's dome with machine learning techniques," *Data Mining and Knowledge Discovery*, vol. 38, no. 3, pp. 1440–1465, 2024.

[20] M. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, pp. 32–64, 1995.

[21] ——, "Measurement of situation awareness in dynamic systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 65–84, 1995.

[22] R. Fang, Q. Zhou, C. Zhou, S. Yuan, K. Wang, Q. Hui, Y. Zhang, and J. Li, "A systematic review of empirical studies on situation awareness: Perspectives from the interaction among humans, machines, and the task environment," *PsyCh Journal*, vol. 14, pp. 718–733, 2025.

[23] M. Endsley, "Situation awareness global assessment technique (sagat)," in *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, 1988, pp. 789–795 vol.3.

[24] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[25] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 135–149, 2012.

[26] P. Nunes, J. Santos, and E. Rocha, "Challenges in predictive maintenance – a review," *CIRP Journal of Manufacturing Science and Technology*, vol. 40, pp. 53–67, 2023.

[27] M. Ravan, M. J. Revez, I. V. Pinto, P. Brum, and J. Birkmann, "A vulnerability assessment framework for cultural heritage sites: The case of the roman ruins of tróia," *International Journal of Disaster Risk Science*, vol. 14, no. 1, p. 26 – 40, 2023.

[28] C. Daly, "A framework for assessing the vulnerability of archaeological sites to climate change: Theory, development, and application," *Conservation and Management of Archaeological Sites*, vol. 16, pp. 268–282, 2014.

[29] R. Figueiredo, X. Romão, and E. Paupério, "Component-based flood vulnerability modelling for cultural heritage buildings," *International Journal of Disaster Risk Reduction*, vol. 61, p. 102323, 2021.

[30] A. Bonazza, A. Sardella, A. Kaiser, R. Cacciotti, P. De Nuntiis, C. Hanus, I. Maxwell, T. Drdácký, and M. Drdácký, "Safeguarding cultural heritage from climate change related hydrometeorological hazards in central europe," *International Journal of Disaster Risk Reduction*, vol. 63, p. 102455, 2021.

[31] P. Brum, F. Neto, M. Revez, I. Pinto, and A. Magalhães, "Icoa883: Storm: Safeguarding cultural heritage through technical and organizational resources management. a european project from a local view," in *19th ICOMOS General Assembly & Scientific Symposium "Heritage and Democracy*, 2017.

[32] S. Della Torre, "Italian perspective on the planned preventive conservation of architectural heritage," *Frontiers of Architectural Research*, vol. 10, no. 1, pp. 108–116, 2021.

[33] G. Acampa and C. M. Parisi, "Management of maintenance costs in cultural heritage," in *Appraisal and Valuation: Contemporary Issues and New Frontiers*, P. Morano, A. Oppio, P. Rosato, L. Sdino, and F. Tajani, Eds. Cham: Springer International Publishing, 2021, pp. 195–212.

[34] R. Cacciotti, A. Sardella, M. Drdácký, and A. Bonazza, "A methodology for vulnerability assessment of cultural heritage in extreme climate changes," *International Journal of Disaster Risk Science*, vol. 15, no. 3, p. 404 – 420, 2024.

[35] A. Roders, "Monitoring cultural significance and impact assessments," in *In Proceedings of the 33rd Annual Meeting of the International Association for Impact Assessment (IAIA13), 13–16 May 2013, Calgary, Alberta, Canada*, 2013.

[36] A. Gandini, A. Egusquiza, L. Garmendia, and J. San-José, "Vulnerability assessment of cultural heritage sites towards flooding events," *IOP Conference Series: Materials Science and Engineering*, vol. 364, p. 012028, 2018.

[37] G. Forino, J. Mackee, and J. von Meding, "A proposed assessment index for climate change-related risk for cultural heritage protection in newcastle (australia)," *International Journal of Disaster Risk Reduction*, vol. 19, 2016.

[38] R. Ortiz and P. Ortiz, "Vulnerability index: A new approach for preventive conservation of monuments," *International Journal of Architectural Heritage*, vol. 10, 2016.

[39] "Unisdr (united nations international strategy for disaster reduction), proposed updated terminology on disaster risk reduction: A technical review," *Geneva: The United Nations Office for Disaster Risk Reduction*, 2015.

[40] R. Pachauri and L. Meyer, "Climate change 2014: Synthesis report." in *Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change, Geneva IPCC*, 2014, p. 117–130.

[41] "Unddr (united nations office for disaster risk reduction), report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction." *Geneva: UNDRR*, 2016.

[42] A. K. Tyagi and N. Sreenath, "Cyber physical systems: Analyses, challenges and possible solutions," *Internet of Things and Cyber-Physical Systems*, vol. 1, pp. 22–33, 2021.

[43] T. Sobb, B. Turnbull, and N. Moustafa, "A holistic review of cyber–physical–social systems: New directions and opportunities," *Sensors*, vol. 23, no. 17, 2023.

[44] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21 980–22 012, 2020.

[45] F. Emmert-Streib, "Defining a digital twin: A data science-based unification," *Machine Learning and Knowledge Extraction*, vol. 5, no. 3, pp. 1036–1054, 2023.

[46] A. Sharma, E. Kosasih, J. Zhang, A. Brintrup, and A. Calinescu, "Digital twins: State of the art theory and practice, challenges, and open research questions," *Journal of Industrial Information Integration*, vol. 30, p. 100383, 2022.

[47] M. Abdelrahman, E. Macatulad, B. Lei, M. Quintana, C. Miller, and F. Biljecki, "What is a digital twin anyway? deriving the definition for the built environment from over 15,000 scientific publications," *Building and Environment*, vol. 274, p. 112748, 2025.

[48] M. Casillo, L. Cecere, F. Colace, A. Lorusso, D. Santaniello, and C. Valentino, "Digital twin and metaverse supporting smart cities: New perspectives and potentials," in *Intelligent Sustainable Systems*, A. K. Nagar, D. S. Jat, D. K. Mishra, and A. Joshi, Eds. Singapore: Springer Nature Singapore, 2024, pp. 111–119.

[49] R. Finelli, P. Sena, A. Lorusso, L. Cecere, and F. Villecco, "Reproduction of road scenarios for simulated driving using lidar surveying technique," *Machines*, vol. 13, no. 1, 2025.

[50] M. Casillo, L. Cecere, F. Colace, A. Lorusso, and D. Santaniello, "Integrating the internet of things (iot) in spa medicine: Innovations and challenges in digital wellness," *Computers*, vol. 13, no. 3, 2024.

[51] Z. B. R. Chavez, A. Lorusso, D. Santaniello, and M. C. De Simone, "A framework for structural monitoring: Deep learning and digital twin for anomaly detection," in *New Technologies, Development and Application VIII*, I. Karabegović, A. Kovačević, and S. Mandžuka, Eds. Cham: Springer Nature Switzerland, 2025, pp. 31–38.

[52] T. J. Emerson, J. M. Reising, and H. G. Britten-Austin, "Workload and situation awareness in future aircraft," *SAE Transactions*, vol. 96, pp. 1130–1137, 1987.

[53] M. A. Dalrymple and S. G. Schiflett, "Measuring situational awareness of awacs weapons directors," in *Situational Awareness in the Tactical Air Environment: Augmented Proceedings of the Naval Air Warfare Center's First Annual Symposium*, 1997.

[54] S. Dusire and C. Mundutéguy, "Intent recognition and situation awareness in transport activities," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 1, pp. 173–176, 2000.

[55] J. Son, Z. Aziz, and F. Peña-Mora, "Supporting disaster response and recovery through improved situation awareness," *Structural Survey*, vol. 26, no. 5, p. 411 – 425, 2008.

[56] C. Chauvin, J. Clostermann, and J.-M. Hoc, "Situation awareness and the decision-making process in a dynamic situation: Avoiding collisions at sea," *Journal of Cognitive Engineering and Decision Making*, vol. 2, no. 1, p. 1 – 23, 2008.

[57] M. Endsley, "Situation awareness misconceptions and misunderstandings," *Journal of Cognitive Engineering and Decision Making*, vol. 9, pp. 4–32, 2015.

[58] ——, "Theoretical underpinnings of situation awareness: A critical review," *Situation awareness analysis and measurement*, pp. 3–32, 2000.

[59] ——, *Designing for Situation Awareness: An Approach to User-Centered Design*, 2nd ed. CRC Press, 2004.

[60] Q. Wang, C. Yang, J. Lu, F. Wu, and R. Xu, "Analysis of preservation priority of historic buildings along the subway based on matter-element model," *Journal of Cultural Heritage*, vol. 45, pp. 291–302, 2020.

[61] A. Lorusso and G. Celenta, "Internet of things in the construction industry: A general overview," in *New Technologies, Development and Application VI*, I. Karabegovic, A. Kovačević, and S. Mandzuka, Eds. Cham: Springer Nature Switzerland, 2023, pp. 577–584.

[62] L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A. S. Albahri, B. S. N. Al-dabbagh, M. A. Fadhel, M. Manoufali, J. Zhang, A. H. Al-Timemy, Y. Duan, A. Abdullah, L. Farhan, Y. Lu, A. Gupta, F. Albu, A. Abbosh, and Y. Gu, "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, p. 46, 2023.

[63] S. Barba, A. di Filippo, M. Limongiello, and B. Messina, "Integration of active sensors for geometric analysis of the chapel of the holy shroud," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W15, pp. 149–156, 2019.

[64] M. Casillo, F. Colace, B. B. Gupta, A. Lorusso, F. Marongiu, and D. Santaniello, "A deep learning approach to protecting cultural heritage buildings through iot-based systems," in *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2022, pp. 252–256.

[65] F. Colace, D. Conte, G. Frasca-Caccia, A. Lorusso, D. Santaniello, and C. Valentino, "An iot-based framework for the enjoyment and protection of cultural heritage artifacts," in *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2023, pp. 489–494.

[66] L. Cecere, F. Colace, A. Lorusso, F. Marongiu, M. Pellegrino, and D. Santaniello, "Iot and deep learning for smart energy management," in *Proceedings of Eighth International Congress on Information and Communication Technology*, X.-S. Yang, R. S. Sherratt, N. Dey, and A. Joshi, Eds. Singapore: Springer Nature Singapore, 2024, pp. 1037–1046.

[67] M. Cosovic, A. Amelio, and E. Junuz, "Classification methods in cultural heritage," in *Proceedings of the 1st International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding (VIPERC 2019), co-located with IRCDL 2019*, ser. CEUR Workshop Proceedings, vol. 2320, Pisa, Italy, 2019, pp. 13–24.

[68] UNESCO, "Recommendation concerning the protection and promotion of museums and collections, their diversity and their role in society," 2015, adopted 17 November 2015, Paris, France.

[69] R. Sizyakin, B. Cornelis, L. Meeus, H. Dubois, M. Martens, V. Voronin, and A. Pizurica, "Crack detection in paintings using convolutional neural networks," *IEEE Access*, vol. PP, pp. 1–1, 2020.

[70] Q. Yuan, X. He, X. Han, and H. Guo, "Automatic recognition of craquelure and paint loss on polychrome paintings of the palace museum using improved u-net," *Heritage Science*, vol. 11, no. 1, p. 65, 2023.

[71] I. Garrido, J. Erazo-Aux, S. Lagüela, S. Sfarra, C. Ibarra-Castanedo, E. Pivarčiová, G. Gargiulo, X. Maldague, and P. Arias, "Introduction of deep learning in thermographic monitoring of cultural heritage and improvement by automatic thermogram pre-processing algorithms," *Sensors*, vol. 21, no. 3, 2021.

[72] Z. Zhang, H. Zhang, J. Hu, S. Sfarra, M. Mostacci, Y. Wang, D. Yang, X. Maldague, D. Niu, and Y. Duan, "Defect detection: An improved yolox network applied to a replica of "the birth of venus" by botticelli," *Journal of Cultural Heritage*, vol. 62, pp. 404–411, 2023.

[73] K. Kavkler, M. Humar, D. Kržišnik, M. Turk, Črtomir Tavzes, C. Gostinčar, S. Džeroski, S. Popov, A. Penko, N. Gunde - Cimerman, and P. Zalar, "A multidisciplinary study of biodeteriorated celje ceiling, a tempera painting on canvas," *International Biodeterioration & Biodegradation*, vol. 170, p. 105389, 2022.

[74] S. Popov, K. Kavkler, and S. Džeroski, "Using machine learning to identify factors contributing to mould in the celje ceiling painting," in *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2021, pp. 217–222.

[75] A. Jalandoni, Y. Zhang, and N. A. Zaidi, "On the use of machine learning methods in rock art research with application to automatic painted rock art identification," *Journal of Archaeological Science*, vol. 144, p. 105629, 2022.

[76] Y. Zeng and Y. Gong, "Nearest neighbor based digital restoration of damaged ancient chinese paintings," in *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, 2018, pp. 1–5.

[77] Z. Zou, P. Zhao, and X. Zhao, "Virtual restoration of the colored paintings on weathered beams in the forbidden city using multiple deep learning algorithms," *Advanced Engineering Informatics*, vol. 50, p. 101421, 2021.

[78] V. Gupta, N. Sambyal, A. Sharma, and P. Kumar, "Restoration of artwork using deep neural networks," *Evolving Systems*, vol. 12, no. 2, pp. 439–446, 2021.

[79] T. Yu, C. Lin, S. Zhang, C. Wang, X. Ding, H. An, X. Liu, T. Qu, L. Wan, S. You, J. Wu, and J. Zhang, "Artificial intelligence for dunhuang cultural heritage protection: The project and the dataset," *International Journal of Computer Vision*, vol. 130, no. 11, pp. 2646–2673, 2022.

[80] I. Nurmala Sari and W. Du, "Structure-texture consistent painting completion for artworks," *IEEE Access*, vol. PP, pp. 1–1, 2023.

[81] B. Sankar, M. Saravanan, K. Kumar, and S. Dubakka, "Transforming pixels into a masterpiece: Ai-powered art restoration using a novel distributed denoising cnn (ddcnn)," in *2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, 2023, pp. 164–175.

[82] A. Bombini, F. G.-A. Bofias, C. Ruberto, and F. Taccetti, "A cloud-native application for digital restoration of cultural heritage using nuclear imaging: Thespian-xrf," *Rendiconti Lincei. Scienze Fisiche e Naturali*, vol. 34, no. 3, pp. 867–887, 2023.

[83] G. Chen, Z. Wen, and F. Hou, "Application of computer image processing technology in old artistic design restoration," *Heliyon*, vol. 9, no. 11, p. e21366, 2023.

[84] P. Kumar, V. Gupta, and M. Grover, "Dual attention and channel transformer based generative adversarial network for restoration of the damaged artwork," *Engineering Applications of Artificial Intelligence*, vol. 128, p. 107457, 2024.

[85] Y. Mei, L. Yang, M. Wang, T. Yu, and K. Wu, "Dunhuangstitch: Unsupervised deep image stitching of dunhuang murals," *IEEE transactions on visualization and computer graphics*, vol. PP, 2024.

[86] G. Sciutto, S. Legrand, E. Catelli, S. Prati, C. Malegori, P. Oliveri, K. Janssens, and R. Mazzeo, "Macroscopic mid-ftir mapping and clustering-based automated data-reduction: An advanced diagnostic tool for in situ investigations of artworks," *Talanta*, vol. 209, p. 120575, 2020.

[87] S. Saha, J. Martusewicz, N. L. W. Streeton, and R. Sitnik, "Segmentation of change in surface geometry analysis for cultural heritage applications," *Sensors*, vol. 21, no. 14, 2021.

[88] A. Kaur, A. Raj, N. Jayanthi, and S. Indu, "Inpainting of irregular holes in a manuscript using unet and partial convolution," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2020, pp. 778–784.

[89] R. Sivan, P. B. Pati, and M. W. A. Kesiman, "Image quality determination of palm leaf heritage documents using integrated discrete cosine transform features with vision transformer," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 28, no. 1, pp. 41–57, 2025.

[90] J. Tao, G. Li, Q. Sun, Y. Chen, D. Xiao, and H. Feng, "An approach for identifying historic village using deep learning," *SN Applied Sciences*, vol. 5, no. 1, p. 13, 2022.

[91] M. Altaweel, A. Khelifi, Z. Li, A. Squitieri, T. Basmaji, and M. Ghazal, "Automated archaeological feature detection using deep learning on optical uav imagery: Preliminary results," *Remote Sensing*, vol. 14, no. 3, 2022.

[92] Øivind Due Trier, J. H. Reksten, and K. Løseth, "Automated mapping of cultural heritage in norway from airborne lidar data using faster r-cnn," *International Journal of Applied Earth Observation and Geoinformation*, vol. 95, p. 102241, 2021.

[93] N. Anttiroiko, F. J. Groesz, J. Ikäheimo, A. Kelloniemi, R. Nurmi, S. Rostad, and O. Seitsonen, "Detecting the archaeological traces of tar production kilns in the northern boreal forests based on airborne laser scanning and deep learning," *Remote Sensing*, vol. 15, no. 7, 2023.

[94] Y. Li, Y. Wang, D. Sui, and M. Guo, "Dense buddha head object detection and counting yolov8 network based on multi-scale attention and data augmentation fusion," *npj Heritage Science*, vol. 13, no. 1, p. 23, 2025.

[95] S. Bhuvaneswari and K. Kathiravan, "Enhancing epigraphy: a deep learning approach to recognize and analyze tamil ancient inscriptions," *Neural Computing and Applications*, vol. 36, no. 31, pp. 19 839–19 861, 2024.

[96] Y. Zhang, Z. Zhang, W. Zhao, and Q. Li, "Crack segmentation on earthen heritage site surfaces," *Applied Sciences*, vol. 12, no. 24, 2022.

[97] M. Altaweel, A. Khelifi, and M. M. Shana'ah, "Monitoring looting at cultural heritage sites: Applying deep learning on optical unmanned aerial vehicles data as a solution," *Soc. Sci. Comput. Rev.*, vol. 42, no. 2, p. 480–495, 2024.

[98] H. Qiu, J. Zhang, L. Zhuo, Q. Xiao, Z. Chen, and H. Tian, "Research on intelligent monitoring technology for roof damage of traditional chinese residential buildings based on improved yolov8: taking ancient villages in southern fujian as an example," *Heritage Science*, vol. 12, no. 1, p. 231, 2024.

[99] L. Peng, W. Bo, H. Yang, and X. Li, "Deep learning-based image compression for enhanced hyperspectral processing in the protection of stone cultural relics," *Expert Systems with Applications*, vol. 271, p. 126691, 2025.

[100] D. Corradetti and J. D. Rodrigues, "Identification of stone deterioration patterns with large multimodal models. definitions and benchmarking," *Journal of Cultural Heritage*, vol. 71, pp. 175–183, 2025.

[101] D. Kwon and J. Yu, "Ravit-ae: Unsupervised anomaly detection for intelligent cultural heritage monitoring using region-attentive vit autoencoder," *IEEE Access*, vol. 12, pp. 180 767–180 780, 2024.

[102] H. Richards-Rissetto, D. Newton, and A. Al Zadjali, "A 3d point cloud deep learning approach using lidar to identify ancient maya archaeological sites," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. VIII-M-1-2021, pp. 133–139, 2021.

[103] A. Mertel, P. Ondrejka, and K. Šabatová, "Spatial predictive modeling of prehistoric sites in the bohemian-moravian highlands based on graph similarity analysis," *Open Geosciences*, vol. 10, no. 1, pp. 261–274, 2018.

[104] I. Wachtel, R. Zidon, S. Garti, and G. Shelach-Lavi, "Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north israel and north-east china," *Journal of Archaeological Science*, vol. 92, pp. 28–36, 2018.

[105] P. Yaworsky, K. Vernon, J. Spangler, S. Brewer, and B. Codding, "Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the grand staircase-escalante national monument," *PLOS ONE*, vol. 15, 2020.

[106] G. Imen, G. Halima, K. Ayoub, and A. Djamel, "Utilizing the maxent machine learning model to forecast urban heritage sites in the desert regions of southwestern algeria: A case study in the saoura region," *Archaeological Prospection*, vol. 31, 2023.

[107] Z. Zou, X. Zhao, P. Zhao, F. Qi, and N. Wang, "Cnn-based statistics and location estimation of missing components in routine inspection of historic buildings," *Journal of Cultural Heritage*, vol. 38, pp. 221–230, 2019.

[108] Y. Liu, M. Hou, A. Li, Y. Dong, L. Xie, and Y. Ji, "Automatic detection of timber-cracks in wooden architectural heritage using yolov3 algorithm," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B2-2020, pp. 1471–1476, 2020.

[109] N. Wang, X. Zhao, Z. Zou, P. Zhao, and F. Qi, "Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, pp. 277–291, 2019.

[110] M. Mishra, T. Barman, and G. V. Ramana, "Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage," *Journal of Civil Structural Health Monitoring*, vol. 14, no. 1, pp. 103–120, 2024.

[111] L. Mansuri and D. Patel, "Artificial intelligence-based automatic visual inspection system for built heritage," *Smart and Sustainable Built Environment*, vol. 11, pp. 622–646, 2021.

[112] E. Elhariri, N. El-Bendary, and S. A. Taie, "Automated pixel-level deep crack segmentation on historical surfaces using u-net models," *Algorithms*, vol. 15, no. 8, 2022.

[113] M. Samhouri, L. Al-Arabiat, and F. Al-Atrash, "Prediction and measurement of damage to architectural heritages facades using convolutional neural networks," *Neural Computing and Applications*, vol. 34, no. 20, pp. 18 125–18 141, 2022.

[114] S.-Y. Lee and H.-H. Cho, "Damage detection and safety diagnosis for immovable cultural assets using deep learning framework," in *2023 25th International Conference on Advanced Communication Technology (ICACT)*, 2023, pp. 310–313.

[115] S. Bruno, R. Galantucci, and A. Musicco, "Decay detection in historic buildings through image-based deep learning," *VITRUVIO - International Journal of Architectural Technology and Sustainability*, vol. 8, 2023.

[116] D. Kwon and J. Yu, "Automatic damage detection of stone cultural property based on deep learning algorithm," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W15, pp. 639–643, 2019.

[117] W.-Y. Choi, J.-W. Park, and S.-Y. Lee, "Gan based deep learning model for detecting damage and displacement of cultural asset," in *2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2021, pp. 1–4.

[118] J. Lee and J. M. Yu, "Automatic surface damage classification developed based on deep learning for wooden architectural heritage," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-M-1-2023, pp. 151–157, 2023.

[119] P. S. Roy, V. Kukreja, V. Jain, and S. Vats, "Classification of defective intensity levels of paint in heritage buildings using the cnn-svm technique," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2023, pp. 17–22.

[120] Sohail, D. Kumar, and V. Kukreja, "Ensemble deep learning using faster rcnn model and fuzzy rule system for health monitoring of heritage castles," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–6.

[121] S. Mehta, V. Kukreja, and A. Gupta, "Exploring the efficacy of cnn and svm models for automated damage severity classification in heritage buildings," in *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 2023, pp. 252–257.

[122] S. Meklati, K. Boussora, M. E. H. Abdi, and S.-A. Berrani, "Surface damage identification for heritage site protection: A mobile crowd-sensing solution based on deep learning," *J. Comput. Cult. Herit.*, vol. 16, no. 2, 2023.

[123] C.-X. Yu, W.-P. Chen, C.-Y. Ju, T.-Y. Chen, K.-C. Li, C.-A. Chen, M.-L. Chan, and S.-L. Chen, "Application of deep learning techniques for thermal imagery analysis in abnormal identification of floor tiles in heritage environments," in *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2023, pp. 1878–1884.

[124] S. Gao, G. Huang, X. Chen, H. Jiang, L. Zhou, and X. Gao, "Two-stage deep learning-based video image recognition of early fires in heritage buildings," *Engineering Applications of Artificial Intelligence*, vol. 129, p. 107598, 2024.

[125] T. Bakirman, B. Kulavuz, and B. Bayram, "Use of artificial intelligence toward climate-neutral cultural heritage," *Photogrammetric Engineering & Remote Sensing*, vol. 89, no. 3, pp. 163–171, 2023.

[126] N. Wang, X. Zhao, L. Wang, and Z. Zou, "Novel system for rapid investigation and damage detection in cultural heritage conservation based on deep learning," *Journal of Infrastructure Systems*, vol. 25, 2019.

[127] A. Lorusso, B. Messina, and D. Santaniello, "The use of generative adversarial network as graphical support for historical urban renovation," in *ICGG 2022 - Proceedings of the 20th International Conference on Geometry and Graphics*, L.-Y. Cheng, Ed. Cham: Springer International Publishing, 2023, pp. 738–748.

[128] M. Cheng, X. Zhang, L. Xia, J. Chen, J. Xie, Z. Ma, and Q. Li, "Visual defect detection for historical building preservation," *Expert Systems with Applications*, vol. 291, p. 128376, 2025.

[129] S. K. Singh, D. Maity, and P. K. Kumawat, "Deep learning-based damage detection and segmentation in the battledore of darbhanga fort," *Journal of Cultural Heritage*, vol. 73, pp. 510–523, 2025.

[130] L. Long, Z. Gan, Z. Liu, B. Zhao, and Q. Li, "Msd-det: Masonry structures damage detection dataset for preventive conservation of heritage," *Journal of Cultural Heritage*, vol. 73, pp. 358–370, 2025.

[131] K. Idjaton, R. Janvier, M. Balawi, X. Desquesnes, X. Brunetaud, and S. Treuillet, "Detection of limestone spalling in 3d survey images using deep learning," *Automation in Construction*, vol. 152, p. 104919, 2023.

[132] A. Ramírez-Arellano, E. M. Muñoz-Silva, M. Antonio-Cruz, and J. I. Vasquez-Gomez, "Deng entropy and lstm neural network to classify built cultural heritage with severe damage," *Journal of Cultural Heritage*, vol. 73, pp. 286–294, 2025.

[133] G. Zonno, R. Aguilar, R. Boroschek, and P. B. Lourenço, "Analysis of the long and short-term effects of temperature and humidity on the structural properties of adobe buildings using continuous monitoring," *Engineering Structures*, vol. 196, p. 109299, 2019.

[134] P. Miglioranza, A. Scanu, G. Simionato, N. Sinigaglia, and A. Califano, "Machine learning and engineering feature approaches to detect events perturbing the indoor microclimate in ringebu and heddal stave churches (norway)," *International Journal of Building Pathology and Adaptation*, vol. 42, 2022.

[135] F. Carrara, F. Falchi, M. Girardi, N. Messina, C. Padovani, and D. Pellegrini, "Deep learning for structural health monitoring: An application to heritage structures," 2023, pp. 581–586.

[136] F. Colace, C. Elia, C. G. Guida, A. Lorusso, F. Marongiu, and D. Santaniello, "An iot-based framework to protect cultural heritage buildings," in *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2021, pp. 377–382.

[137] M. Casillo, C. G. Guida, M. Lombardi, A. Lorusso, F. Marongiu, and D. Santaniello, "Predictive preservation of historic buildings through iot-based system," in *2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON)*, 2022, pp. 1194–1198.

[138] M. Casillo, F. Colace, R. Gaeta, A. Lorusso, D. Santaniello, and C. Valentino, "Revolutionizing cultural heritage preservation: an innovative iot-based framework for protecting historical buildings," *Evolutionary Intelligence*, vol. 17, no. 5, pp. 3815–3831, 2024.

[139] G. Nota and G. Petraglia, "Heritage buildings management: the role of situational awareness and cyber-physical systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 15, no. 4, pp. 2227–2239, 2024.

[140] F. Granata and F. Di Nunno, "Artificial intelligence models for prediction of the tide level in venice," *Stochastic Environmental Research and Risk Assessment*, vol. 35, no. 12, pp. 2537–2548, 2021.

[141] A. Cardellicchio, S. Ruggieri, A. Nettis, V. Renò, and G. Uva, "Physical interpretation of machine learning-based recognition of defects for the risk management of existing bridge heritage," *Engineering Failure Analysis*, vol. 149, p. 107237, 2023.

[142] W. Lee and D.-h. Lee, "Cultural heritage and the intelligent internet of things," *J. Comput. Cult. Herit.*, vol. 12, no. 3, 2019.

[143] F. Piccialli, F. Giampaolo, G. Casolla, V. S. D. Cola, and K. Li, "A deep learning approach for path prediction in a location-based iot system," *Pervasive and Mobile Computing*, vol. 66, p. 101210, 2020.

[144] C. Martella, A. Miraglia, J. Frost, M. Cattani, and M. van Steen, "Visualizing, clustering, and predicting the behavior of museum visitors," *Pervasive and Mobile Computing*, vol. 38, pp. 430–443, 2017, special Issue IEEE International Conference on Pervasive Computing and Communications (PerCom) 2016.

[145] M. Della Lucia, G. Dore, and R. M. Umar, "Handling the open culture dilemma in museum management: an exploratory interdisciplinary study," *Scientometrics*, vol. 129, no. 12, pp. 7699–7733, 2024.

[146] L. Cao and J. Yin, "Research on sharing behavior strategy of cultural heritage institutions based on evolutionary game theory," *Sustainability*, vol. 15, no. 13, 2023.

[147] S. Tomasiello and Z. Alijani, "Fuzzy-based approaches for agri-food supply chains: a mini-review," *Soft Computing*, vol. 25, no. 11, pp. 7479–7492, 2021.

[148] S. Tomasiello, W. Pedrycz, and V. Loia, "On fractional tikhonov regularization: Application to the adaptive network-based fuzzy inference system for regression problems," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 11, pp. 4717–4727, 2022.

[149] R. Loffredo and M. De Santo, "Using ontologies for LLM applications in cultural heritage," in *CEUR Workshop Proceedings*, vol. 3865, 2024, p. 37 – 43.

[150] G. D'Aniello, M. De Falco, M. Gaeta, and M. Lepore, "A situation-aware learning system based on fuzzy cognitive maps to increase learner motivation and engagement," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2020, pp. 1–8.

[151] G. D'Aniello, A. Gaeta, V. Loia, and F. Orciuoli, "Integrating gso and saw ontologies to enable situation awareness in green fleet management," in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2016, pp. 138–144.

[152] M. Mishra and P. B. Lourenço, "Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review," *Journal of Cultural Heritage*, vol. 66, pp. 536–550, 2024.

[153] O. Gómez-Carmona, D. Casado-Mansilla, D. López-de Ipiña, and J. García-Zubia, "Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches," *Internet of Things (Netherlands)*, vol. 25, 2024.

[154] S. H. Alsamhi, S. Kumar, A. Hawbani, A. V. Shvetsov, L. Zhao, and M. Guizani, "Synergy of human-centered ai and cyber-physical-social systems for enhanced cognitive situation awareness: Applications, challenges and opportunities," *Cognitive Computation*, vol. 16, no. 5, p. 2735 – 2755, 2024.

[155] L. Fan, D. Cao, C. Zeng, B. Li, Y. Li, and F.-Y. Wang, "Cognitive-based crack detection for road maintenance: An integrated system in cyber-physical-social systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 6, pp. 3485–3500, 2023.

[156] C. Meng, X. S. Jiang, X. M. Wei, and T. Wei, "A time convolutional network based outlier detection for multidimensional time series in cyber-physical-social systems," *IEEE Access*, vol. 8, p. 74933 – 74942, 2020.

[157] L. Faramondi, F. Flammini, S. Guarino, and R. Setola, "A hybrid behavior- and bayesian network-based framework for cyber–physical anomaly detection," *Computers and Electrical Engineering*, vol. 112, 2023.

[158] F. Rodrigues, V. Cotella, H. Rodrigues, E. Rocha, F. Freitas, and R. Matos, "Application of deep learning approach for the classification of buildings' degradation state in a bim methodology," *Applied Sciences (Switzerland)*, vol. 12, no. 15, 2022.

[159] L. Cecere, F. Colace, A. Lorusso, and D. Santaniello, "Predictive maintenance of an archeological park: An iot and digital twin based approach," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14735 LNAI, p. 323 – 341, 2024.

[160] K. Rohit, P. Singh, N. Patel, P. Kamat, and S. Kumar, "Efficient aerial object detection: An exploration with yolov8," in *Machine Learning Algorithms*, M. Khurana, A. Thakur, P. Kantha, C.-S. Shieh, and R. K. Shukla, Eds. Cham: Springer Nature Switzerland, 2025, pp. 360–371.

[161] G. D'Aniello, M. Gaeta, R. Gravina, Q. Li, Z. U. Rehman, and G. Fortino, "Situation identification in smart wearable computing systems based on machine learning and context space theory," *Information Fusion*, vol. 104, 2024.

[162] L. Aliberti, G. D'Aniello, G. Fortino, and M. Gaeta, "Situation projection for enhanced human-machine interaction based on rule mining," in *2024 IEEE 4th International Conference on Human-Machine Systems, ICHMS 2024*, 2024.

[163] A. Padovitz, S. Loke, and A. Zaslavsky, "Towards a theory of context spaces," in *IEEE Annual Conference on Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second*, 2004, pp. 38 – 42.

[164] A. Marsico, V. De Santis, and D. Capolongo, "Erosion rate of the aliano biancana badlands based on a 3d multi-temporal high-resolution survey and implications for wind-driven rain," *Land*, vol. 10, p. 828, 2021.

[165] R. Monjo, "Measure of rainfall time structure using the dimensionless n-index," *Climate Research*, vol. 67, 2016.

[166] T. Beer, "Beaufort wind scale," in *Encyclopedia of Earth Sciences Series*, 2013, pp. 42–45.

[167] M. N. Mistry, R. Schneider, P. Masselot, D. Royé, B. Armstrong *et al.*, "Comparison of weather station and climate reanalysis data for modelling temperature-related mortality," *Scientific Reports*, vol. 12, no. 1, p. 5178, 2022.

[168] G. Gualtieri, "Reliability of era5 reanalysis data for wind resource assessment: A comparison against tall towers," *Energies*, vol. 14, p. 4169, 2021.

[169] Q. Jiang, W. Li, Z. Fan, X. He, W. Sun, S. Chen, J. Wen, J. Gao, and J. Wang, "Evaluation of the era5 reanalysis precipitation dataset over chinese mainland," *Journal of Hydrology*, vol. 595, p. 125660, 2021.

[170] P. Wang, T. Hu, F. Gao, R. Wu, W. Guo, and X. Zhu, "A hybrid data-driven framework for spatiotemporal traffic flow data imputation," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16 343–16 352, 2022.

[171] D. Adhikari, W. Jiang, J. Zhan, M. Assefa, H. A. Khorshidi, U. Aickelin, and D. B. Rawat, "A lightweight-window-portion-based multiple imputation for extreme missing gaps in iot systems," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 3676–3689, 2024.

[172] M. C. De Simone, A. Lorusso, and D. Santaniello, "Predictive maintenance and structural health monitoring via iot system," 2022, 2022 IEEE Workshop on Complexity in Engineering, COMPENG 2022.

[173] H. Ge, M. Matsui, and N. Koshizuka, "An open-iot approach on elevator for enabling autonomous robotic vertical mobility," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 2021, pp. 139–141.

[174] J. Zhang, M. Huang, N. Wan, Z. Deng, Z. He, and J. Luo, "Missing measurement data recovery methods in structural health monitoring: The state, challenges and case study," *Measurement: Journal of the International Measurement Confederation*, vol. 231, 2024.

[175] G. D'Aniello, M. Gaeta, and T.-P. Hong, "Effective quality-aware sensor data management," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 65–77, 2018.

[176] E. Bellini, G. D'Aniello, F. Flammini, and R. Gaeta, "Situation awareness for cyber resilience: A review," *International Journal of Critical Infrastructure Protection*, vol. 49, p. 100755, 2025.

[177] W. Powell, M. Foth, S. Cao, and V. Natanelov, "Garbage in garbage out: The precarious link between iot and blockchain in food supply chains," *Journal of Industrial Information Integration*, vol. 25, 2022.

[178] X. Wu, M. Xu, J. Fang, and X. Wu, "A multi-attention tensor completion network for spatiotemporal traffic data imputation," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20 203–20 213, 2022.

[179] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Missing value imputation for industrial iot sensor data with large gaps," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6855–6867, 2020.

[180] D. A. Guastella, G. Marcillaud, and C. Valenti, "Edge-based missing data imputation in large-scale environments," *Information*, vol. 12, no. 5, 2021.

[181] Y. Shen, D. Zhao, X. Hu, W. Pedrycz, Y. Chen, J. Li, and Z. Xiao, "Structure identification of missing data: a perspective from granular computing," *Soft Computing*, 2024.

[182] X. Hu, Y. Shen, W. Pedrycz, Y. Li, and G. Wu, "Granular fuzzy rule-based modeling with incomplete data representation," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, p. 6420 – 6433, 2022.

[183] X. Hu, W. Pedrycz, K. Wu, and Y. Shen, "Information granule-based classifier: A development of granular imputation of missing data," *Knowledge-Based Systems*, vol. 214, 2021.

[184] H. Nugroho, N. P. Utama, and K. Surendro, "Comparison method for handling missing data in clinical studies," in *Proceedings of the 2020 9th International Conference on Software and Computer Applications*, ser. ICSCA '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 46–50.

[185] Y. Bian, G. Y. Yi, and W. He, "A unified framework of analyzing missing data and variable selection using regularized likelihood," *Computational Statistics and Data Analysis*, vol. 194, 2024.

[186] Y. Sun, J. Li, Y. Xu, T. Zhang, and X. Wang, "Deep learning versus conventional methods for missing data imputation: A review and comparative study," *Expert Systems with Applications*, vol. 227, 2023.

[187] J. Li, S. Guo, R. Ma, J. He, X. Zhang, D. Rui, Y. Ding, Y. Li, L. Jian, J. Cheng, and H. Guo, "Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets," *BMC Medical Research Methodology*, vol. 24, no. 1, 2024.

[188] R. C. Pereira, P. H. Abreu, P. P. Rodrigues, and M. A. Figueiredo, "Imputation of data missing not at random: Artificial generation and benchmark analysis," *Expert Systems with Applications*, vol. 249, 2024.

[189] Q. Song and M. Shepperd, "Missing data imputation techniques," *Int. J. Bus. Intell. Data Min.*, vol. 2, no. 3, p. 261–291, 2007.

[190] D. P. Anil Jadhav and K. Ramanathan, "Comparison of performance of data imputation methods for numeric dataset," *Applied Artificial Intelligence*, vol. 33, no. 10, pp. 913–933, 2019.

[191] S. Annas, P. Kartikasari, and R. Arisandi, "Handling incomplete data with regression imputation," *Journal of Physics: Conference Series*, vol. 1752, no. 1, p. 012049, 2021.

[192] O. Giustolisi, D. Savic, and A. Doglioni, "Data reconstruction and forecasting by evolutionary polynomial regression," in *Hydroinformatics*, 2004, pp. 1245–1252.

[193] S. Arciniegas-Alarcón, M. García-Peña, W. J. Krzanowski, and C. Rengifo, "Missing value imputation in a data matrix using the regularised singular value decomposition," *MethodsX*, vol. 11, p. 102289, 2023.

[194] S. Faisal and G. Tutz, "Multiple imputation using nearest neighbor methods," *Information Sciences*, vol. 570, pp. 500–516, 2021.

[195] S. Philip, P. Vashisth, A. Chaturvedi, and N. Gupta, "Imputation of missing values using improved k-means clustering algorithm to attain data quality," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2021, pp. 295–301.

[196] F. H. Mausor, J. Jaafar, and S. M. Taib, "Missing values imputation using fuzzy c means based on correlation of variable," in *2020 International Conference on Computational Intelligence (ICCI)*, 2020, pp. 261–265.

[197] J. Dongre, G. L. Prajapati, and S. V. Tokekar, "The role of apriori algorithm for finding the association rules in data mining," in *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014, pp. 657–660.

[198] H. K. D. Sarma and S. Mishra, "Mining time series data with apriori tid algorithm," in *2016 International Conference on Information Technology (ICIT)*, 2016, pp. 160–164.

[199] K. R. Thakre and R. Shende, "Implementation on an approach for mining of datasets using apriori hybrid algorithm," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 2017, pp. 939–943.

[200] C. Borgelt, "An implementation of the fp-growth algorithm," in *Proceedings of the 1st International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations*, ser. OSDM '05. New York, NY, USA: Association for Computing Machinery, 2005, p. 1–5.

[201] J. Wu, Q. Song, and J. Shen, "An novel association rule mining based missing nominal data imputation method," in *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007)*, vol. 3, 2007, pp. 244–249.

[202] ——, "Missing nominal data imputation using association rule based on weighted voting method," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, pp. 1157–1162.

[203] N. Jiang, "A data imputation model in sensor databases," in *High Performance Computing and Communications*, R. Perrott, B. M. Chapman, J. Subhlok, R. F. de Mello, and L. T. Yang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 86–96.

[204] A. Chaurasia, "Combining rules for f- and beta-statistics from multiply-imputed data," *Econometrics and Statistics*, vol. 25, pp. 51–65, 2023.

[205] T.-P. Hong and C.-W. Wu, "Mining rules from an incomplete dataset with a high missing rate," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3931–3936, 2011.

[206] J. Yoon, W. Zame, and M. Schaar, "Estimating missing data in temporal data streams using multi-directional recurrent neural networks," *IEEE Transactions on Biomedical Engineering*, vol. PP, 2017.

[207] W. Du, D. Côté, and Y. Liu, "Saits: Self-attention-based imputation for time series," *Expert Systems with Applications*, vol. 219, p. 119619, 2023.

[208] W. Pedrycz, "Granular computing - the emerging paradigm," *Jour. of Uncer. Sys.*, vol. 1, 2001.

[209] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1977–1989, 2013.

[210] L.-J. Li, M.-Z. Li, and J.-S. Mi, "Granular structure evaluation and selection based on justifiable granularity principle," *Information Sciences*, vol. 665, 2024.

[211] L. Zadeh, "Key roles of information granulation and fuzzy logic in human reasoning, concept formulation and computing with words," in *Proceedings of IEEE 5th International Fuzzy Systems*, vol. 1, 1996, pp. 1–1.

[212] Y. Yao, "Interval sets and interval-set algebras," in *2009 8th IEEE International Conference on Cognitive Informatics*, 2009, pp. 307–314.

[213] N. Dăneţ, "Interval analysis - a powerfull trend in numerical analysis," in *International Conference "Trends and Challenges in Applied Mathematics" - ICTCAM 2007*, 2007.

[214] C. Zhong, W. Pedrycz, D. Wang, L. Li, and Z. Li, "Granular data imputation: A framework of granular computing," *Applied Soft Computing*, vol. 46, pp. 307–316, 2016.

[215] A. Ragel and B. Crémilleux, "Treatment of missing values for association rules," in *Research and Development in Knowledge Discovery and Data Mining, Second Pacific-Asia Conference, PAKDD-98, Melbourne, Australia, April 15-17, 1998, Proceedings*, ser. Lecture Notes in Computer Science, X. Wu, K. Ramamohanarao, and K. B. Korb, Eds., vol. 1394. Springer, 1998, pp. 258–270.

[216] A. Ragel and B. Crémilleux, "Mvc—a preprocessing method to deal with missing values," *Knowledge-Based Systems*, vol. 12, no. 5, pp. 285–291, 1999.

[217] X. He, F. Min, and W. Zhu, "Comparison of discretization approaches for granular association rule mining," *Canadian Journal of Electrical and Computer Engineering*, vol. 37, no. 3, pp. 157–167, 2014.

[218] W. Pedrycz, "Granular computing: an introduction," in *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*, vol. 3, 2001, pp. 1349–1354 vol.3.

[219] M. Biba, F. Esposito, S. Ferilli, N. Di Mauro, and T. M. A. Basile, "Unsupervised discretization using kernel density estimation," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, ser. IJCAI'07. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, p. 696–701.

[220] S. Zhang, B. Guo, A. Dong, J. He, Z. Xu, and S. X. Chen, "Cautionary tales on air-quality improvement in beijing," *Proceedings. Mathematical, physical, and engineering sciences*, vol. 473, no. 2205, p. 20170457, 2017.

[221] B. Agbo, H. Al-Aqrabi, R. Hill, and T. Alsboui, "Missing data imputation in the internet of things sensor networks," *Future Internet*, vol. 14, no. 5, 2022.

[222] N. U. Okafor and D. T. Delaney, "Missing data imputation on iot sensor networks: Implications for on-site sensor calibration," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 22 833–22 845, 2021.

[223] S. Jäger, A. Allhorn, and F. Biessmann, "A benchmark for data imputation methods," *Frontiers in Big Data*, vol. 4, 2021.

[224] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, 2020, pp. 10 778–10 787.

[225] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning (ICML)*, 2019.

[226] Y. Cai, T. Luan, H. Gao, H. Wang, L. Chen, Y. Li, M. A. Sotelo, and Z. Li, "Yolov4-5d: An effective and efficient object detector for autonomous driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[227] J. Li, C. Liu, X. Lu, and B. Wu, "Cme-yolov5: An efficient object detection network for densely spaced fish and small targets," *Water*, vol. 14, no. 15, 2022.

[228] P. Dvořák, J. Müllerová, T. Bartaloš, and J. Brůna, "Unmanned aerial vehicles for alien plant species detection and monitoring," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W4, pp. 83–90, 2015.

[229] D. Yadav, M. Choksi, and M. A. Zaveri, "Supervised learning based greenery region detection using unnamed aerial vehicle for smart city application," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2019, pp. 1–7.

[230] Q. Wang, M. Cheng, X. Xiao, H. Yuan, J. Zhu, C. Fan, and J. Zhang, "An image segmentation method based on deep learning for damage assessment of the invasive weed solanum rostratum dunal," *Computers and Electronics in Agriculture*, vol. 188, p. 106320, 2021.

[231] D. A. Yudin, V. Adeshkin, A. V. Dolzhenko, A. Polyakov, and A. E. Naumov, "Roof defect segmentation on aerial images using neural networks," in *Advances in Neural Computation, Machine Learning, and Cognitive Research IV*, B. Kryzhanovsky, W. Dunin-Barkowski, V. Redko, and Y. Tiumentsev, Eds. Cham: Springer International Publishing, 2021, pp. 175–183.

[232] Y. Pi, N. D. Nath, and A. H. Behzadan, "Detection and semantic segmentation of disaster damage in uav footage," *Journal of Computing in Civil Engineering*, vol. 35, no. 2, p. 04020063, 2021.

[233] Y. Jing, Y. Ren, Y. Liu, D. Wang, and L. Yu, "Automatic extraction of damaged houses by earthquake based on improved yolov5: A case study in yangbi," *Remote Sensing*, vol. 14, no. 2, 2022.

[234] Y. Pi, N. D. Nath, and A. H. Behzadan, "Convolutional neural networks for object detection in aerial imagery for disaster response and recovery," *Advanced Engineering Informatics*, vol. 43, p. 101009, 2020.

[235] Y. Wang, W. Feng, K. Jiang, Q. Li, R. Lv, and J. Tu, "Real-time damaged building region detection based on improved yolov5s and embedded system from uav images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4205–4217, 2023.

[236] M. Christaki, C. Vasilakos, E.-E. Papadopoulou, G. Tataris, I. Siarkos, and N. Soulakellis, "Building change detection based on a gray-level co-occurrence matrix and artificial neural networks," *Drones*, vol. 6, no. 12, 2022.

[237] J. Bin, R. Zhang, R. Wang, Y. Cao, Y. Zheng, E. Blasch, and Z. Liu, "An efficient and uncertainty-aware decision support system for disaster response using aerial imagery," *Sensors*, vol. 22, no. 19, 2022.

[238] Q. D. Cao and Y. Choe, "Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks," *Natural Hazards*, vol. 103, no. 3, p. 3357–3376, 2020.

[239] G. Wang, S.-Y. Shin, and G. Jo, "An improved post-hurricane building damaged detection method based on transfer learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, p. 1546, 2024.

[240] A. Ishraq, A. A. Lima, M. M. Kabir, M. S. Rahman, and M. F. Mridha, "Assessment of building damage on post-hurricane satellite imagery using improved cnn," in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, 2022, pp. 665–669.

[241] M. Böge, D. Bulatov, and L. Lucks, "Localization and grading of building roof damages in high-resolution aerial images," in *Computer Vision, Imaging and Computer Graphics Theory and Applications*, A. P. Cláudio, K. Bouatouch, M. Chessa, A. Paljic, A. Kerren, C. Hurter, A. Tremeau, and G. M. Farinella, Eds. Cham: Springer International Publishing, 2020, pp. 497–519.

[242] N. Kumari, P. Moral, D. Mustafi, A. Mustafi, S. Anwar, and S. K. Sahana, "Deep learning for post-hurricane damage detection with sar-based analysis using densenet201 and svm," in *15th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2024*, vol. 2, 2024, p. 1349 – 1358.

[243] A. Ramesh, S. K. R. Prasad, S. Srikanth, and S. Tripathi, "Hurricane damage detection using computer vision," in *ACM International Conference Proceeding Series*, 2023, p. 126 – 132.

[244] Z. Zhang, "A high-accuracy ensemble method of convolutional neural networks for damaged building detection on post hurricane satellite images," in *CAIBDA 2022; 2nd International Conference on Artificial Intelligence, Big Data and Algorithms*, 2022, pp. 1–4.

[245] M. Żarski and J. A. Miszczak, "Multi-step feature fusion for natural disaster damage assessment on satellite images," *IEEE Access*, vol. 12, pp. 140 072–140 081, 2024.

[246] Z. Hong, H. Zhong, H. Pan, J. Liu, R. Zhou, Y. Zhang, Y. Han, J. Wang, S. Yang, and C. Zhong, "Classification of building damage using a novel convolutional neural network based on post-disaster aerial images," *Sensors*, vol. 22, no. 15, 2022.

[247] T. Rajaperumal and C. C. Columbus, "Transforming the electrical grid: the role of ai in advancing smart, sustainable, and secure energy systems," *Energy Informatics*, vol. 8, no. 1, 2025.

[248] Y. Yue, G. Yan, T. Lan, R. Cao, Q. Gao, W. Gao, B. Huang, G. Huang, Z. Huang, Z. Kan, X. Li, D. Liu, X. Liu, D. Ma, L. Wang, J. Xia, X. Yang, M. Zhou, A. G.-O. Yeh, R. Guo, and C. Claramunt, "Shaping future sustainable cities with ai-powered urban informatics: Toward human-ai symbiosis," *Computational Urban Science*, vol. 5, no. 1, 2025.

[249] A. Tabbakh, L. Al Amin, M. Islam, G. M. I. Mahmud, I. K. Chowdhury, and M. S. H. Mukta, "Towards sustainable ai: a comprehensive framework for green ai," *Discover Sustainability*, vol. 5, no. 1, 2024.

[250] I. Salehin, M. S. Islam, P. Saha, S. Noman, A. Tuni, M. M. Hasan, and M. A. Baten, "Automl: A systematic review on automated machine learning with neural architecture search," *Journal of Information and Intelligence*, vol. 2, no. 1, pp. 52–81, 2024.

[251] R. S. Olson and J. H. Moore, "Tpot: A tree-based pipeline optimization tool for automating machine learning," in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds. Cham: Springer International Publishing, 2019, pp. 151–160.

[252] K. T. Chitty-Venkata, M. Emani, V. Vishwanath, and A. K. Somani, "Neural architecture search benchmarks: Insights and survey," *IEEE Access*, vol. 11, pp. 25 217–25 236, 2023.

[253] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-weka: combined selection and hyperparameter optimization of classification algorithms," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 847–855.

[254] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[255] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.

[256] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–7475.

[257] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using yolo: challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[258] Z. Qiu, S. Rong, and L. Ye, "Yolf-shippnet: Improved retinanet with pyramid vision transformer," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 58, 2023.

[259] N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, and A. Smola, "Autogluon-tabular: Robust and accurate automl for structured data," *arXiv:2003.06505*, 2020.

[260] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in *Proceedings of the Twenty-First International Conference on Machine Learning*, ser. ICML '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 18.

[261] M. Feurer, A. Klein, K. Eggensperger, J. T. Springenberg, M. Blum, and F. Hutter, "Auto-sklearn: Efficient and robust automated machine learning," in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds. Cham: Springer International Publishing, 2019, pp. 113–134.

[262] K. Solmann, R. Loffredo, and S. Tomasiello, "Using fuzzy transforms for neural networks-based wireless localization in outdoor environments," *Neural Computing and Applications*, vol. 36, no. 32, pp. 20 027–20 041, 2024.

[263] S. Tomasiello, "Least-squares fuzzy transforms and autoencoders: Some remarks and application," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 129–136, 2021.

[264] T. M. H. Gedara, V. Loia, and S. Tomasiello, "Using fuzzy transform for sustainable fake news detection," *Applied Soft Computing*, vol. 151, p. 111173, 2024.

[265] G. Sun, Z. Jiao, A. Zhang, F. Li, H. Fu, and Z. Li, "Hyperspectral image-based vegetation index (hsvi): A new vegetation index for urban ecological research," *International Journal of Applied Earth Observation and Geoinformation*, vol. 103, p. 102529, 2021.

[266] B. Mensah, N. Rai, K. Betitame, and X. Sun, "Advances in weed identification using hyperspectral imaging: A comprehensive review of platform sensors and deep learning techniques," *Journal of Agriculture and Food Research*, vol. 18, p. 101388, 2024.

[267] R. Taylor, "Situational awareness rating technique (sart): The development of a tool for aircrew systems design," in *Situational Awareness*, 2017, pp. 111–128.

[268] R. J. Brachman and H. J. Levesque, *Knowledge Representation and Reasoning*, ser. The Morgan Kaufmann Series in Artificial Intelligence. Morgan Kaufmann, 2004.

[269] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, 1993.

[270] S. Bloehdorn, P. Cimiano, and A. Hotho, "Learning ontologies to improve text clustering and classification," in *From Data and Information Analysis to Knowledge Engineering*, M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nurnberger,

and W. Gaul, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 334–341.

[271] S. Sbai, M. R. Chbihi Louhdi, H. Behja, and R. Chakhmoune, "Jsontoonto: Building owl2 ontologies from json documents," *International Journal of Advanced Computer Science and Applications*, vol. 10, p. 213, 2019.

[272] L. F. Abbott and P. Dayan, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, ser. Computational Neuroscience Series. The MIT Press, 2001.

[273] C. Koch, *Biophysics of Computation: Information Processing in Single Neurons (Computational Neuroscience Series)*. USA: Oxford University Press, Inc., 2004.

[274] G. Buzsáki, *Rhythms of the Brain*. New York: Oxford University Press, 2006.

[275] G. Norman, "Likert scales, levels of measurement adn the "laws" of statistics," *Advances in health sciences education : theory and practice*, vol. 15, pp. 625–32, 2010.

[276] O. Langsrud, "Anova for unbalanced data: Use type ii instead of type iii sums of squares," *Statistics and Computing*, vol. 13, pp. 163–167, 2003.

[277] D. C. Montgomery, *Design and Analysis of Experiments*, 7th ed. Hoboken, NJ: John Wiley & Sons, 2010.

[278] S. E. Maxwell and H. D. Delaney, *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2004.

[279] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.