Published in: MULTIMEDIA TOOLS AND APPLICATIONS journal.
This version of the article has been accepted for publication, after peer review and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/s11042-022-14290-0

Noname manuscript No.
(will be inserted by the editor)

Emotion Detection for Supporting Depression Screening

Rita Francese · Pasquale Attanasio

Received: date / Accepted: date

Abstract Depression is the most prevalent mental disorder in the world. One of the most adopted tools for depression screening is the Beck Depression Inventory-II (BDI-II) questionnaire. Patients may minimize or exaggerate their answers. Thus, to further examine the patient's mood while filling in the questionnaire, we propose a mobile application that captures the BDI-II patient's responses together with their images and speech. Deep learning techniques such as Convolutional Neural Networks analyze the patient's audio and image data. The application displays the correlation between the patient's emotional scores and DBI-II scores to the clinician at the end of the questionnaire, indicating the relationship between the patient's emotional state and the depression screening score. We conducted a preliminary evaluation involving clinicians and patients to assess (i) the acceptability of proposed application for use in clinics and (ii) the patient user experience. The participants were eight clinicians who tried the tool with 21 of their patients. The results seem to confirm the acceptability of the app in clinical practice.

Keywords emotion detection, depression, user study

1 Introduction

Depression is the most common mental disorder in the world. Worldwide, it is estimated that more than 280 million people suffer from depression[1] and this

Rita Francese*
Department of Computer Science
University of Salerno
E-mail: francese@unisa.it

Pasquale Attanasio
IBM
Naples
E-mail: pasquale_attanasio@it.ibm.com

[1] https://www.who.int/news-room/fact-sheets/detail/depression

number continues to rise. In addition, only in the United States depression causes 200 million lost workdays each year, at the cost of 17 to 44 billion dollars per year[2].

The symptoms of depression severely impact people's quality of life, how they feel, what they think, and how they go about everyday activities. It interferes in relationships with other people or with the way one works or sleeps. It is characterized by sadness and loss of interest and concentration. In the worst cases, a person may be tempted to commit suicide, the second leading cause of death among 15-29 years old[3]. Due to the severity of this illness, it is essential to diagnose it when the first symptoms appear. Often, depressed people are not properly diagnosed. When depression evolves uncontrollably, it may become a severe illness that deeply affects everything a person does in his/her life.

Usually, psychologists or psychiatrists perform depression screening by talking to patients and evaluating their responses. Making a diagnosis is difficult because depressed people can show many different symptoms. They may be apathetic or sad, or drastically reduce food consumption. Clinicians often use standard questionnaires for screening depression. One of the most commonly used validated instruments is the Beck Depression Inventory-II (BDI-II) questionnaire [25], used as screening tool together with clinical observation. Many physiological studies have shown that the speech and facial activities of depressed people have some differences compared to non-depressed people [11] [19]. This fact has been exploited to perform automatic depression screening by analyzing patient emotions [37] [8] [29]. We believe that the experience of the expert matters. Thus, instead of performing automatic depression screening we started investigating how to assist the physician in a non-invasive way: the respondent to a screening questionnaire may exaggerate or minimize the answers. Patients express their emotions while filling out a screening questionnaire, which the physician's mobile device can capture and interpret together with the results of the questionnaire using artificial intelligence tools.

In [12], we proposed a mobile application aiming at capturing images and the voice of the patient while he/she is filling out a screening questionnaire. The system provided the physician with the correlation between the questionnaire scores and the emotion recognition performed by Neural Networks running on the clinician's device that detect the patient's emotional state by analyzing images and speech. A preliminary evaluation was conducted to assess the acceptability of the tool in clinical use, and we have only reflected the clinicians' view. This paper extends our previous work [12] as follows:

- the assessment was improved by including the patient's perspective;
- the background and related work sections were improved;

[2] `https://www.cdc.gov/workplacehealthpromotion/health-strategies/depression/evaluation-measures/index.html`

[3] https://www3.paho.org/spc-crb/index.php?option=com$_c$$ontentview = articleid = 470 : world - health - day - 2017 - qdepression - let - s - talkqItemid = 1540

Table 1: Interpretation of the BDI-II questionnaire scores.

BDI-II Score	Depression Degree
0-13	None
14-19	Mild
20-28	Moderate
29-63	Severe

– the threats to validity, outcomes and limitations of our research were discussed.

This paper is organized as follows: Section 2 summarizes the background concepts on depression screening, emotion representation models, and related work. Section 3 introduces the proposed approach named EDApp (Emotion and Depression Application). The preliminary evaluation of EDApp is described in Section 4, while Section 5 discusses our findings and limitations. Conclusions and future work are traced in Section 6.

2 Background

This section summarizes the main concepts related to the Beck Depression Inventory-II questionnaire, the different approaches to emotion recognition, and the research related to applying machine learning techniques to support depression screening.

2.1 The Beck Depression Inventory-II Questionnaire

There are several self-report instruments in the literature for assessing depression. One of the most commonly used is the Beck Depression Inventory (BDI) questionnaire, adopted in more than 7,000 studies. The BDI is a self-report inventory that measures characteristic attitudes and symptoms of depression. Aaron T. Beck et al. [4] proposed their questionnaire in 1961. Since then, it has been revised twice, in 1978 as BDI-IA [2] and then in 1996 as BDI-II [3]. Originally written in English, the questionnaire has been translated into 17 languages and used in Europe, Asia, and South America.

The questionnaire aims at assessing the severity of depression in individuals aged 13 years and older. It consists of 21 items associated with a depressive content, such as sadness or weight loss.

Item scores range from 0 (none) to 3 (severe). The questionnaire result is obtained by summing up the scores of the items. It is interpreted according to Table 1, which shows the relationships between the BDI-II scores and the level of depression.

BDI-II is generally used as an indicator of depression severity, including "no depression", as shown in Table 1. It suffers from the problems of all self-reports, such as that the respondent may not be sincere in completing the questionnaire. Thus, it is used as screening tool, but together with clinical

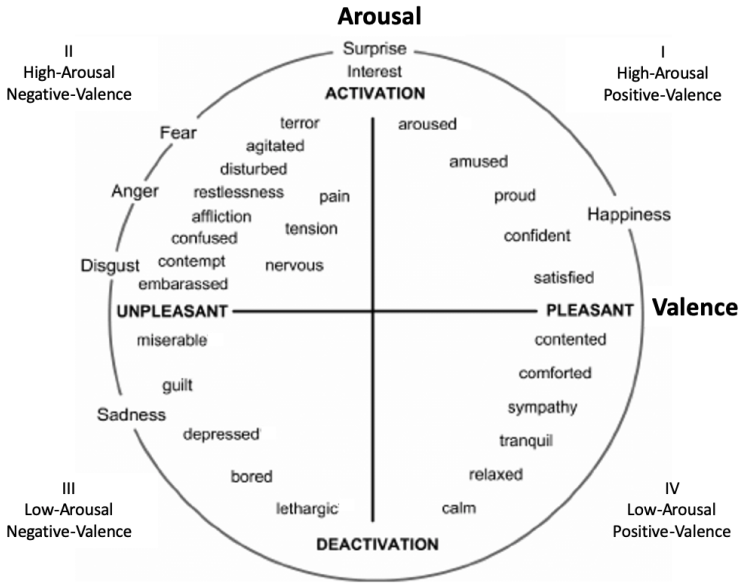


Fig. 1: The Circumflex model of emotions [34].

observation. It is in this direction that EDApp can help: it may reveal a discrepancy between the self-report questionnaire results and the user's mood.

2.2 Emotion representations

In general, emotions are represented in the literature in two different ways: discrete or dimensional. The first approach assumes that humans experience some basic emotions, such as anger, fear, sadness, and happiness, which are associated with physiological responses and vary in intensity [9].

In contrast, the dimensional approach proposed by Russel, called the circumflex model, represents emotions in a two-dimensional space defined by valence and arousal, which lie between Pleasant and Unpleasant and high versus low activation [34]. This model arranges the various emotional states in this two-dimensional space. The circumflex model is shown in Fig. 1.

2.3 Related work

Several research efforts have addressed depression detection by applying machine learning techniques. Many types of data sources have been considered. Some of them are based on handwriting and drawing [22] because depression affects them. Facial images have also been widely studied [32][46] [38], as the facial expressions of people with depression differ from those of healthy people. People's speech is also influenced by depression [40]. Many studies were also

based on physiological data obtained using sensors, such as heart rate [13] and EEG [30]. Audio/video content was often combined with the BDI-II scale [31] to assess the level of depression. Related works were classified according to the following criteria and are listed in Table 2.

- *Year.* Date of publication.
- *Media type.* Types of media used to detect depression, such as audio (A), video (V), and questionnaire (Q).
- *Dataset.* The selected datasets for training the models. Datasets related to depression are very limited due to privacy issues [45]. The available datasets are often inconsistent, e.g., they have different languages, types of subjects, and data types. Many of the related works conducted the experiments using the AVEC2013 and AVEC2014 depression sub-challenge datasets, which provided the estimated depression level of the participants using the Beck Depression Inventory. The language used is German. These two datasets are useful for assessing depression levels. Another relevant dataset is FER2013, which is useful for assessing emotions.
- *Model.* The AI model used.
- *End-users.* Y/N - Did the authors evaluate end-user's experience?
- *Aim.* The goal of the approach:
  - D - Detection. The tool automatically detects depression without the clinician intervention.
  - S - Support. The tool provides the clinician with an indication with information useful to assess depression.

In [18], Jan et al. employed deep learning methods to analyze audio and video and predicted BDI-II score using a linear regression model. Different pre-trained CNN models were tried, including a VGG-Face CNN to capture facial expressions. The dataset used was AVEC2014 [42] challenge, which collects videos of depressed people in German.

In [17], Beck depression level was measured by analyzing the patient's speech through a CNN in combination with hand-crafted features.

In [6], distribution learning was used to search for relationships between the user's images and his/her depression scores. For this purpose, a new expectation loss was proposed to represent the depression distributions.

In [31], the authors proposed the Spatio-Temporal Attention (STA) network and a Feature Fusion (MAFF) strategy to represent the depression cues. A Support Vector Regression Predictor was used to predict the individual depression level on the BDI-II scale.

Mulay et al. [29] predicted depression by considering both the scores of the BDI-II questionnaire and the images extracted from a video of the user. They used the FER2013 dataset and employed a CNN model that achieved 66.45% of accuracy. All six emotions captured by FER2013 were used. Audio was not included. The system receives as input the BDI-II questionnaire, the video of the user, and demographic data and provides a classification of the depression level.

Table 2: Related work

Ref.	Year	Media Type	Dataset	Model	end-users	Aim
[18]	2017	A,V	AVEC2014	Linear Regression, VGG-Face	N	D
[17]	2018	A	AVEC2013, AVEC2014	CNN	N	D
[6]	2019	V	AVEC2013, AVEC2014	ResNet-50	N	D
[31]	2020	A+V	AVEC2013, AVEC2014	STA, MAFF, Support Vector Regression	N	D
[29]	2020	A	FER2013	CNN	N	D
[38]	2021	V	Cohn-Kanade, ad-hoc	SVN	N	D
[36]	2021	A	ad-hoc	Random Forest	N	D
[23]	2021	A	AVEC2013, AVEC2014	attentional residual network	N	D
[43]	2021	A	ad-hoc	SVM	N	D
[14]	2022	A	AVEC2014	Gradient Boosting	N	D
This paper	2021	A,V,Q	FER2013, Demos	CNN	Y	S

Tadalagi and Joshi propose a system for automatic detection of depression [38]. The image is preprocessed using Viola-Jones face recognition algorithm; classification is performed using support vector machine (SVM) and a linear binary pattern descriptor is applied. The Cohn-Kanade dataset was selected for the emotions joy and disgust, while images of contempt were collected from the Internet. These images were used to train the SVM classifier. The accuracy is 80%.

Shi et al. [36] analyzed the audio track of an ad-hoc created dataset starting from 66 subjects. They extracted specific audio-features, such as the average of Zerocrossing Rate, Energy, Entropy of Energy, Spectral Centroid, and Spectral Spread. Random Forest (RF) reached a mean F1 of 0.71. BDI-II score has been adopted to label the training dataset. Also He et al. [23] used BDI-II scores for verifying their depression prediction when applying attentional residual network on Videos of the AVEC2013 and AVEC2014 datasets. The model also estimated the severity of depression.

Audio related to spontaneous speech narratives of patients classified as depressed by using observation and DBI-II questionnaire results was adopted in [43] to classify depressed and healthy patients. An SVM classificator was

trained on four acustic features (Jitter, MFCC, derivatives of cepstral coefficients, and spectral centroid) reaching 85.25% of Accuracy.

Hamiditabar et al. [14] combined acoustic-space and score-space features to estimate Beck's Depression Index (BDI-II). Different machine learning methods were experimented and the best performance was reached by Gradient Boosting.

Unlike most other work, we do not aim to detect depression, but to help the clinician interpret the BDI-II score. Our idea is to determine if the emotional state of the user is related to the answer provided. For this purpose, we perform a correlation analysis between the emotion recognition scores (audio and video) and the scores of the BDI-II items.

3 The proposed approach

We propose an application, named EDApp, i.e. Emotion and Depression Application, which aims at helping an expert interpret the results of the BDI-II questionnaire. EDApp collects nonverbal cues each time the patient answers a question and analyzes the correlation between the emotions detected in the patient's audio and video and the questionnaire results. Correlations provide insight into the emotional state of the patient when completing the questionnaire.

EDApp is a mobile application that runs on the clinicians' smartphone. It was developed using Flutter, Google's user interface toolkit for building multiplatform applications[4].

EDApp collects images and audio recordings of the patient as she fills in the questionnaire on the clinician's device by following the process in Fig. 2. It starts with the patient filling out a consent form on the device. If she agrees, the questionnaire begins. For each BDI-II question, we introduced the request to give a vocal explanation about the given answer. The answer is recorded and evaluated. When the user completes a question, EDApp runs the following subprocesses:

- *image-based emotion detection.* As shown in the middle part of Fig. 2, the application captures an image when the patient terminates to fill in an answer of the questionnaire. This image is then inputted to a CNN that classifies the patient's emotional state.
- *speech-based emotion detection.* The patient's speech is preprocessed: The audio signal is converted into a spectrogram, which is used as input to a CNN that is designed to recognize the user's emotional state in her voice, as in the upper subprocess of Fig. 2.

The answers to the questionnaire are stored together with the emotions detected in the user's speech and images. When she finishes filling out the questionnaire, EDApp calculates the correlation between the distribution of the questionnaire scores and the emotional states detected by audio and video.

[4] https://flutter.dev/

EDApp then generates a report showing the BDI-II responses and scores, the results of the emotional analysis, and the correlation results. In addition, the clinician may examine the user's images and speech.

3.1 The EDApp interaction modality

The EDApp interface proposes the user to answer a question of the BDI-II questionnaire, as in Fig. 3 (a), where a question about sadness is proposed. Then, the app requires him/her to motivate the answer, as in Fig 3(b), while the rightmost interface records the answer. A voice-controlled chatbot was also developed using Dialogflow[5] for accessibility for the blind.

3.2 Image-based emotion recognition

We used the FER2013[6] dataset as the basis for creating the image dataset to train our system. It considers a discrete set of emotions for classifying the images: anger, disgust, fear, happiness, sadness, surprise, and Neutral; it contains 28,709 images. We compiled our dataset by selecting the 1248 sad

[5] https://cloud.google.com/dialogflow

[6] `https://www.kaggle.com/deadskull7/fer2013##__sid=js0`

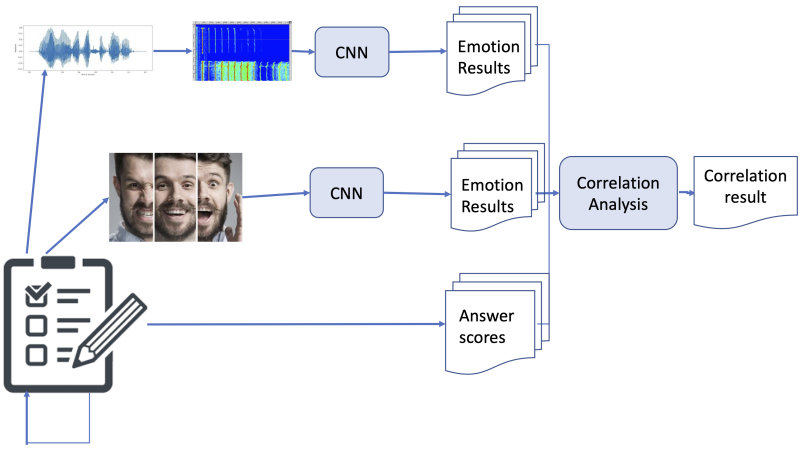
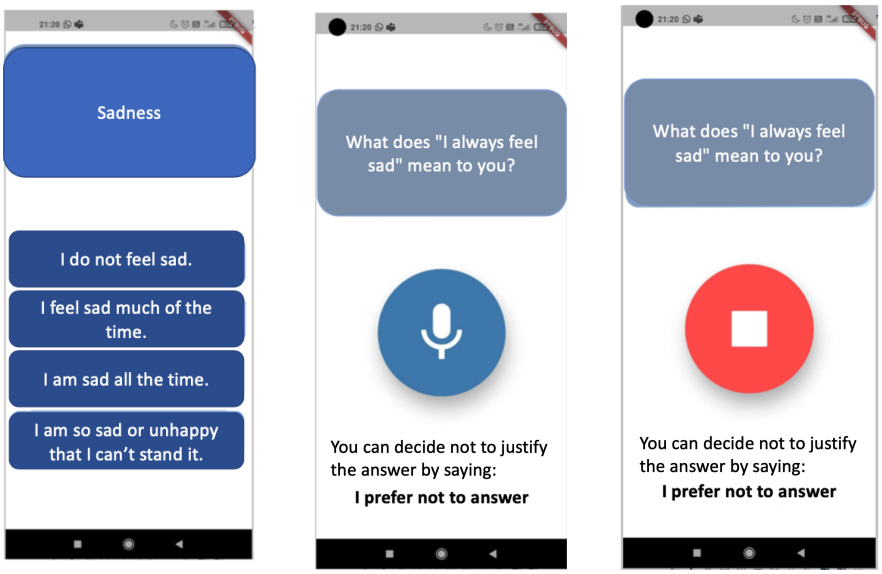


Fig. 2: The process for supporting depression screening performed by EDApp.



(a) (b) (c)

Fig. 3: Some screenshots of the EDApp interface.

images from FER2013 and randomly selected 250 images from each of the other emotions. In this way, we obtained a dataset in which the images are classified as Sad/notSad. This choice was because sadness is the emotion best representing depression [27].

We adopted a Convolutional Neural Network (CNN) model to classify emotions, which is widely used for emotion recognition.

A CNN has a variable number of learning filters. The first level filters enable the detection of simple patterns. As the number of levels increases, the complexity of the recognized patterns increases too. The final levels, which consist of fully connected neurons, are capable of making predictions. The selected CNN model is structured as follows: three convolutional blocks (consisting of Convolutional 2D and Maxpooling layers) and two fully connected (FC) blocks, the last one performing the classification. We trained the model with 100 epochs and a batch size of 50. The Adam optimizer was used. The training set, the test set, and the validation set were determined considering the percentages 80%, 20% of the original dataset, and 20% of the training set, respectively.

All convolutional levels are activated using Rectified Linear Units (ReLUs), while for the last node, we used the sigmoid activation function, which presses all values between 0 and 1 into the shape of a sigmoid curve which has the property of a probability. After the first fully connected block, a dropout level of 0.20 was added. The CNN provides as output a percentage value associated with the detection of the emotion sadness. The output of the sigmoid function enables also us to take a binary decision, $decision(x)$ e.g. y=1 - sad , y= 0 - notSad, where x is the input image.

Table 3: Image and speech emotion classification results.

Classification approach	Accuracy	Sensitivity	Specificity
Image	79.7%	90.2%	72.0%
Speech	71.8%	83.2%	68.7%

$$decision(x) = \begin{cases} 1 \;\; if \; P(y=1|x) > 0.5 \\ 0 \;\; otherwise \end{cases}$$

As shown in Table 3, the classifier based on images achieved an accuracy of 79.7%.

3.3 Speech-based emotion recognition

Machine learning has been adopted to recognize people's emotions by analyzing their speech. Some approaches examine the original audio recording [5], others analyze spectrograms [24]. In this paper, we performed emotional speech analysis based on spectrograms, which represent the strength of a signal in a colored graph chart in terms of Time, Frequency, and Color. Time is associated with the X axis, frequency to the Y axis, and color represents the amplitude of the frequency components at a particular time. A dark color represents a low signal amplitude. A spectrogram is obtained from the audio signal by applying the Fast Fourier transform (FFT).

EDApp is addressed to the Italian People. For this reason, we adopted the Demos [33] dataset, a database collecting Italian emotional audio consisting in 9,365 emotional and 332 neutral speeches produced by 68 Italian people (23 females, 45 males) classified in the following emotions: anger, sadness, happiness, fear, surprise, disgust, and an additional emotion, guilt. Similarly to the previous section, we created a dataset composed of sad and not sad speech. In particular, we selected all the 1,530 speeches labeled sadness and randomly chose 255 samples for each one of the other emotions, except guilt.

We adopted a CNN also in the case of emotional analysis from speech and followed the structure described in [1]. The CNN receives as input a 256 x 256 spectrogram derived from speech audio. Then it is processed by three consecutive convolutional blocks and three fully connected (FC) blocks, the last devoted to providing the classification. Each convolutional level has ReLU as activation function and a max-pooling layer. The last fully-connected layer consists of one node, activated by the sigmoid activation function that will squeeze all the values between 0 and 1 into the form of a sigmoid curve. Dropout layers follow the first two FC layers with 50% as dropout ratio. Dropout layers aim at avoiding overfitting. The training set, test set and validation set have been set by considering the percentages 80%, 20% of the initial dataset, and 20% of the training set, respectively. The training process runs for 30 epochs with a batch size set to 100. The learning rate was initially set to 0.01 with a decay of 0.1 after every ten epochs. We got 71.8% of Accuracy (see Table 3).

Table 4: Correlation Intervals.

Correlation Intervals	**Strength of the correlation**
0.00 to 0.19 (-0.19 to 0)	very weak positive (negative)
0.20 to 0.39 (-0.39 to -0.20)	weak positive (negative)
0.40 to 0.69 (-0.69 to -0.40)	moderate positive (negative)
0.70 to 0.89 (-0.89 to -0.70)	strong positive (negative)
0.90 to 1 (-1 to -0.90)	very strong positive (negative)

3.4 Correlation Analysis

We used correlation analysis to examine the relationships between the results of the BDI-II questionnaire and the results of emotional image recognition. At the same time, we also analyzed the relationship between the results of the BDI-II questionnaire and emotional speech recognition.

We decided to use Spearman's correlation coefficient to measure the correlation since we cannot assume that the sample is normally distributed [16][28]. The Spearman test yields two values: the rho coefficient and a p-value.

Rho ranges from 1 (maximum positive correlation) to -1 (maximum negative correlation). The closer the rho value is to zero, the weaker the correlation, while the closer it is to -1 or 1, the stronger the correlation (Table 4).

The p-value denotes the likelihood that the observed correlation is due to chance. It ranges between 0 and 1. If a p-value is close to 0, the observed correlation is unlikely to be due to chance, and there is a very high probability that the null hypothesis (i.e., there is no correlation between variables) is wrong. In this case (i.e., there is a correlation between the variables), the alternative hypothesis must be accepted. In the interpretation of Spearman's test, a p-value less than or equal to an alpha fixed to 0.05 (5%) is considered statistically significant. Only in the case of $p-value < 0.05$ the coefficient rho may be considered.

We report the correlation results to the clinician by adopting the following three presentation approaches, one numerical and two graphical. The clinician may visualize all the reports or select one of them as default visualization modality.

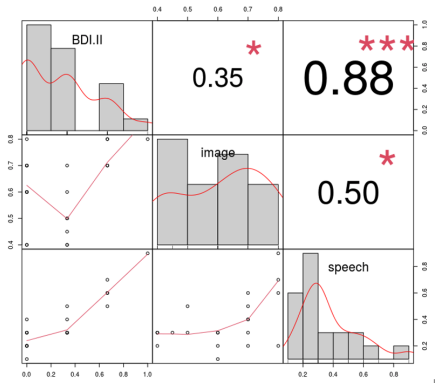
- *Correlation matrix.* The correlation matrix is a table reporting the correlation coefficients between the variables, the BDI-II scores, the image-based emotional analysis, and the speech-based emotional analysis. Each cell of the table contains a correlation coefficient. See, for example, the first three rows of Table 5. The last three rows of the table provide the associated p-values. In this example all the correlations are valid: for all of them $p-value < 0.05$. In particular, the correlation between Speech and BDI-II is strongly positive, while for Image and BDI-II is weakly positive. The correlation between image and speech is moderate. This may be due to difficulty of the mobile device of taking perfect pictures (e.g., cut portion of the person face, see Section 5). The strong correlation of speech and BDI-II scores gives an indication of the relationships between the emotion

Table 5: An example of correlation matrix between BDI-II scores, image and speech emotion detection with the associated p-values.

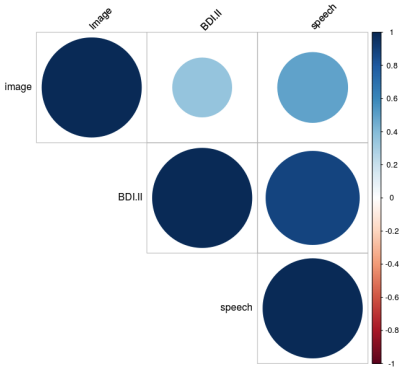
Correlation coefficients			
	BDI-II	*Image*	*Speech*
BDI-II	1.00	0.35	0.88
Image	0.35	1.00	0.50
Speech	0.88	0.50	1.00
p-values			
	BDI-II	*Image*	*Speech*
BDI-II		0.0234	0.0000
Image	0.0234		0.0215
Speech	0.000	0.0215	

in the voice and the answer provided. A weak positive correlation with the images taken from the camera also provides an indication in this direction.

– *Scatter plots.* A scatterplot displays the relationship between two numeric variables by using dots. The values of one of the two variables are shown on the X-axis, and the values of the other one on the Y-axis. A point represents the datum of each subject on the chart. An example related to the correlation results in Table 5 is shown in Fig. 4a. It shows the correlation matrix of the three variables, where the upper triangular of the matrix contains the correlation coefficient values. Stronger correlations are characterized by a heavier and bigger font. Along the diagonal, a histogram shows the distribution of the variables. As an example, the histogram in the first row shows the distribution of the BDI-II normalized score of the considered patient: 9 questions scored 0, 7 of them 0,33, 4 scored 0,66 and the other scored 1. The lower triangular of the matrix represents the bivariate scatterplots, with a fitted line. Each point on the plot shows the X and Y scores for the considered patient. As an example, the plot in the second row, first column, shows the points associated to the variable X=BDI-II scores and Y= image sadness scores related to the same question. Some points overlapped.
  The significance of the correlation is represented by the number of asterisks as follows: * indicates $0.05 < p - value < 0.01$, ** indicates $0.01 < p - value < 0.001$, *** indicates $p - value < 0.001$.
– *Correlograms.* They are a quick way to call attention to the most correlated variables in a dataset. For example, Fig. 4b highlights the strong correlation between BDI-II scores and speech emotional classification. Also, the color has a meaning. Colors towards the blue represent direct correlation, while colors toward the red are associated with indirect correlation. In case the correlation is unsignificant, the cell is marked "X."



(a)



(b)

Fig. 4: Correlation visualization modalities.

3.5 Risk/Privacy and safety requirements

In the last years, the diffusion of mobile devices for accessing to health services has dramatically increased [41]. At the same time, also the indisputable human right to access to health care grows together with tangible ethical concerns of other human rights. The World Health Organization[7] and the Code of Ethics of the European Community[8] identified the following relevant aspects: autonomy and respect of person, beneficence, nonmaleficence, safety, and justice. In the following we detail how EDApp satisfies these requirements.

– *Autonomy and respect of person.* It can be pursued through the use of open-source platforms by providing high value to the requirement of user-friendliness. The end-users should be involved in all the development life-cycle, always adopting informed consent (with all building components),

[7] `https://www.who.int/ethics/Ethics_basic_concepts_ENG.pdf`
[8] https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/CodeofEthicsinAcademicResearch.pdf

and understanding and ownership (e.g., through self-governance). The end-users should be informed on the type of data collected by the application and the implications related to the data collection. In particular, data can be collected only if the users provide a consensus form and they understand what they are consenting to[9].
When systems collect passive data, as data provided by GPS and accelerometer, ethic concerns are particularly relevant because the daily activities of the user may be tracked [20]. EDApp does not track the patient because all the patient data, including multimedia contents, are taken while the questionnaire is filled in and stored on the clinician device without being transmitted. Patients are aware of this from the consensus form. The application has been developed by adopting a user-centered approach for satisfying the user-friendless requirement. The EDApp prototype is not open-source at the present.

- *Beneficence.* The aim of the application is to "do good" for others. This requirement is satisfied by EDApp, whose aim is to support the clinician in the depression screening.
- *Nonmaleficence.* The application has to avoid damaging others.
  The use of EDApp does not cause any risk for the patient health. The results of the correlation are handled by the clinicians.
- *Safety.* It is recommended that privacy and data security should be pursued from the beginning of the app development and in all the development lifecycle phases, respecting the need for evidence-based processes, clear and well-understood rules, procedures, and standards of care [20]. Modern technology such as encryption should be used and the risk of security breaches should be assessed. The development should follow the approach Privacy by Design that guarantees the security constraints from the beginning.
  In the present version, EDApp collects the patient data on the clinician's mobile device on which the computation is performed. Data are not transmitted.
- *Justice.* All the citizens have equal right of being able to access mHealth applications. To grant this right, an mHealth application should be easy-to-use, robust, interoperable with the other healthcare systems, and accessible to all citizens. Data should be rectifiable, portable on a different platform, and erasable.
  We concentrate our attention on the need to provide an easy-to-use application for non-expert users. We considered also accessibility for blind people: to support accessibility of impaired vision people the app has been equipped by a chatbot that supports the user interaction and reads the BDI-II questions. EDApp identifies ownership; it is a research product experimented by the University of Salerno. It has been developed without funding, with the work of master students. It provides support to the clinician and does not claim to perform screening. The app works offline and

[9] `https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm`

runs on Android and Ios operating systems because it has been developed with a multiplatform framework.

4 A preliminary evaluation

4.1 Goal

The goal of this preliminary evaluation is to answer the following Research Question: **RQ:** May EDApp be adopted in clinical use?

To answer this question we have to consider both the clinician's and the patient's point of view.

Therefore, we detailed our RQ as follows:

- **RQ1:** Is the support offered by EDApp acceptable in routine clinician use?
- **RQ2:** Is the User experience adequate from the patient point-of-view?

4.2 Design

EDApp was adopted during the clinicans' screening activity from January 7 2021 to February 10 2021. We conducted a prospective acceptability study [21], designed to assess both clinicians' acceptability of EDApp in clinical use and the patients' user experience.

The experiment tasks performed by the patients consisted in using EDApp for filling in the BDI-II questionnaire, while the clinicians' task consisted in examining the report provided by the tool on the relationship between the patient's emotions and the BDI-II scores, for each one of their patients. A report consists in the correlation analysis results, as those provided in Table 5 and Fig. 4. The perception of the clinicians on the use of the tool in the routine clinic is collected by a questionnaire. The engagement of the patients is measured by administrating them the User Experience Questionnaire (UEQ), often adopted in the literature in case of special needs people with mental health problems (e.g., [7]), or specifically with depression (e.g., [26]).

4.2.1 Participants

This preliminary study involved both clinicians and their patients. Participants provided written informed consent, in which they declared that they (1) were informed about the study details, (2) understood what the research involved, (3) understood what their consent was needed for; (4) might refuse to participate in the research at any time during the research project; (5) had the opportunity to ask the experimenter questions and receive answers to those questions. Finally, patients agreed to be recorded during the study by using the clinician device, and consented to the processing of their demographic data we received in anonymous form to the extent necessary for the implementation of the research project, including the data collected with the questionnaires

Table 6: Participants' age.

Group	Mean	St.dev	Min	Max
Clinician	43	6.18	35	55
Patient	36.24	14.02	19	62

Table 7: On average, how often do you use your WWW browser (for a specific set of tasks or activities)?

	Clinicians	Patients
More than 9 times/day	5	3
5 to 8 times/day	2	3
1 to 4 times/day	1	7
A few times a week	-	5
Once a week	-	1
Once a month or less	-	2

Table 8: On average, how many hours a week do you use your WWW browser?

	Clinicians	Patients
0 to 1 hours/week	-	1
2 to 4 hours/week	1	7
5 to 6 hours/week	2	3
7 to 9 hours/week	4	4
10 to 20 hours/week	1	5
21 to 40 hours/week	-	1
Over 40 hours/week	-	-

in anonymous form. Clinicians and patients filled in a demographic questionnaire. The purpose is to get information on their age, their familiarity with the use of the smartphone, of the personal computer and Internet. Questions in Table 7, 8, and 9 are extracted from the Web and Internet usage questionnaire[10]. Further details on the two categories of participants are reported in the following.

– *Clinicians.* Clinicians were four psychiatrists and four psychologists of Campania region, Italy, who provided voluntary and written informed consent. Five of them were female.
Most clinicians were between the ages of 35-45, as shown in Table 6. Regarding the technological skills of clinicians, most of them were regular users of technology: as shown in Fig. 5, all of them used a smartphone for chatting, most of them for playing and capturing audio/video, three rarely installed new applications. Only one clinician connected to the Internet one to 4 times a day for performing tasks. 7 of them used the browser for at least 5 hours at week and the personal computer at least 7 hours/week.

[10] `https://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/questions/use.html`

Table 9: On average, how many hours per week do you use your personal computer?

	Clinicians	Patients
0 to 1 hours/week	-	-
2 to 4 hours/week	1	8
5 to 6 hours/week	-	3
7 to 9 hours/week	4	5
10 to 20 hours/week	3	4
21 to 40 hours/week	-	1
Over 40 hours/week	-	-

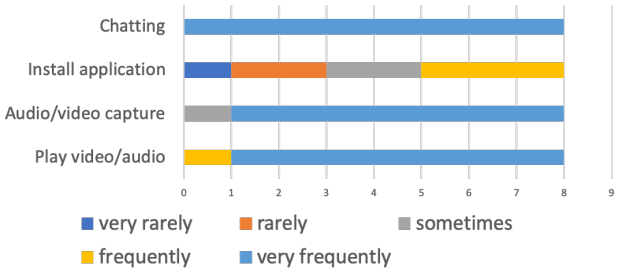


Fig. 5: Clinicians' use of smartphone.

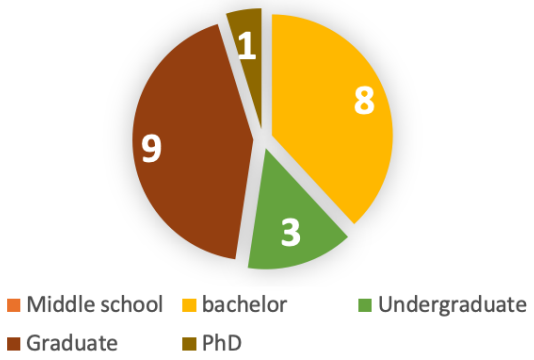


Fig. 6: Educational level of patients.

– *Patients.* Clinicians involved 21 adult patients in the study. 13 of them were female. Concerning their educational level, as shown in Fig. 6 they got all at least the bachelor level.
55% of the patients answered frequently and very frequently in all the scales related to the use of smartphone questions, see Fig. 7. In particular, 85%

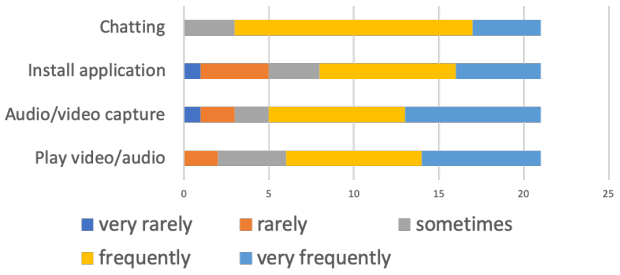


Fig. 7: Patients' use of smartphone.

frequently used it for chatting. We also asked for the frequency of use of the browser and 90,48% of them used the browser at least once a day for performing a task (not only for surfing). 61% used their personal computer at least five hours at week.

4.2.2 Apparatus and materials

Clinicians'acceptability is relevant for their intention to use EDApp in the clinical routine (RQ1). Concerning the patients, user experience may impact on their feelings, impressions and attitudes. It is a measure of how comfortable a patient is during the interaction (RQ2).

Thus, we evaluated the acceptability of our approach by requiring the clinicians to fill a questionnaire in which the first three questions are extracted from the Acceptability E-scale questionnaire proposed by Tariman et al. [39]. The questionnaire for the clinicians is shown in Table 10. It consists of seven closed questions with 5-points Likert Scale and an open question (Q8) for collecting opinions and suggestions. We extended the original version related to the acceptability of a generic tool for e-health with specific questions on the correlation metaphors (Q4) and the satisfaction of the clinician on the information provided by the correlation analysis that compares the result of the BDI-II score with the patient's depression as derived from pictures and/or speech (Q5). We also asked the clinician's opinion on the use of EDApp in the routine practice (Q6) and on her overall satisfaction (Q7).

We also adopted the Net Promoter Score (NPS) as an indicator of the clinician loyalty towards the product [15]. It consists of the following question measured on a scale from 0 to 10: *how likely are you to recommend EDApp to a colleague?"*. Respondents that score from 0 to 6 are called 'detractors', participants that score from 9 to 10 are called 'promoters'. NPS is computed by subtracting the percentage of 'detractors' from the percentage of 'promoters.' It ranges between -100 (worst) and +100 (best).

Patients filled in the UEQ questionnaire, consisting in the following six scales with 26 items [35]:

Table 10: Acceptability Questionnaire.

ID	Question
Q1	How easy was EDApp for you to use? Scale: 1 - very difficult, 5 - very easy
Q2	How helpful was the emotional analysis offered by EDApp in supporting the depression detection? Scale: 1 - very unhelpful, 5 - very helpful
Q3	Was the amount of time EDApp took to complete acceptable? Scale: 1 - very unacceptable, 5 -very acceptable
Q4	Are the correlation representations easy to understand? Scale: 1- difficult to understand, 5 - easy to understand
Q5	Are you satisfied with the correlation results w.r.t. the BDI-II data filled in by your patients? Scale: 1 - very dissatisfied, 5 - very satisfied
Q6	How would you rate your willingness to use the EDApp application for routine practice ? Scale: 1- never, 5 - every time
Q7	How would you rate your overall satisfaction with EDApp? Scale: 1 - very dissatisfied, 5 - very satisfied
Q8	Please, detail your opinion and suggestions on the tool (open)

- Attractiveness: it assesses the user's overall impression of the product.
- Perspicuity: it concerns the learnability of the tool and the easiness to get familiar with it.
- Efficiency: it assesses if the users can perform the tasks without unnecessary effort.
- Dependability: it concerns the user's perception to control the interaction.
- Stimulation: it assesses if the product is stimulating and motivates the users in using it.
- Novelty: it concerns the innovativeness and creativeness of the tool and if it can capture users' interest.

A benchmark is associated with UEQ [35] aiming at providing a tool to test if a product has sufficient user experience. This benchmark lets us compare the results of UEQ with the ones of other products of a dataset containing quite different typical products.

4.3 Procedure

We adopted the following procedure:

- one of the authors of this paper presented EDApp aims and features to the clinicians and how it manages personal data (as described in Section 3.5). This activity was conducted in about fifteen minutes in distance modality because of covid-19 pandemic restrictions;
- clinicians filled in an active consensus form and a demographic questionnaire;
- clinicians selected patients and proposed to use EDApp, available on the clinician's mobile device. They also informed patients that their videos and speech, together with the DBI-II scores, would be stored on the clinician's device, see Section 3.5. In particular, patients were informed of the study goal by the clinicians before starting the test. They filled in an online consensus form, which also detailed the data management.

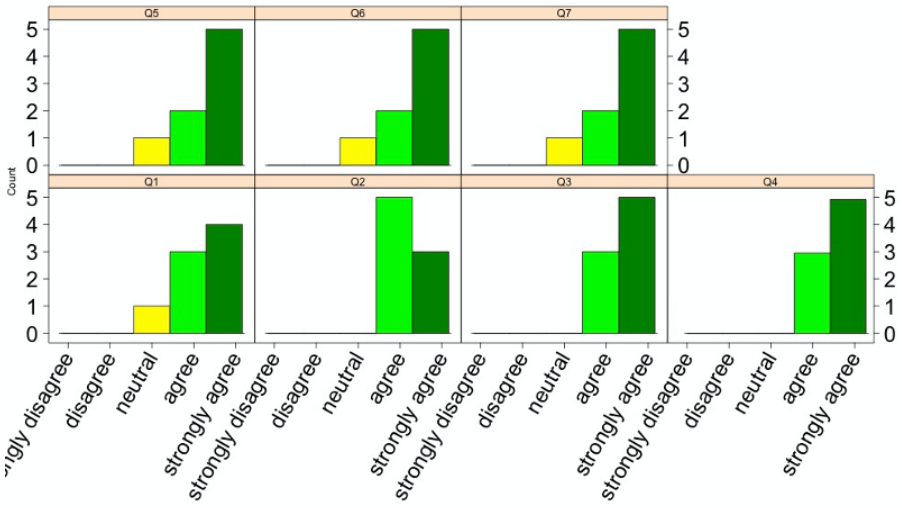


Fig. 8: The results of the Acceptability Questionnaire.

- patients accepted all the conditions, provided a consent form, and filled in the BDI-II questionnaire by using EDApp;
- each patient filled-in the UEQ questionnaire on the clinician device;
- clinicians examined the results on their device;
- at the end of the experimentation, clinicians filled in the Acceptability Questionnaire and the NPS question and sent the UEQ questionnaire results of all their patients to the researcher in anonymous form.

4.4 Results

In this section we report the results of the evaluation related to the two research questions.

4.4.1 RQ1: Is the support offered by EDApp acceptable in routine clinician use?

The histograms depicted in Fig. 8 summarizes the answers the clinicians provided by filling in the Acceptability Questionnaire. In particular, seven of them positively considered EDApp in terms of easiness to use (**Q1**). The neutral clinician had low familiarity with the use of the PC. All of them judged emotional analysis feature helpful for supporting depression screening (**Q2**). EDApp also required an acceptable time to accomplish the questionnaire (**Q3**). The different correlation representations were judged understandable by all of them (**Q4**); this may be because correlation analysis techniques is largely used by

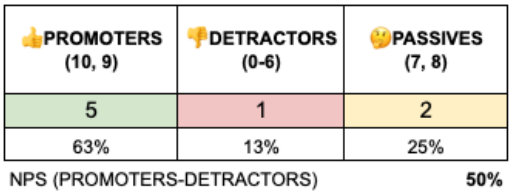


Fig. 9: The EDApp NPS score.

Table 11: UEQ statistics.

Scale	Mean	St.dev
Attractiveness	1.667	1.16
Perspicuity	1.909	0.89
Efficiency	1.995	0.74
Dependability	1.239	0.44
Stimulation	1.159	1.01
Novelty	1.682	0.50

clinicians. Seven of them considered satisfying the results provided by the correlation analysis; one was neutral (**Q5**). Seven clinicians would use EDApp in the clinical practice (**Q6**). Concerning the global satisfaction, five clinicians were very satisfied of EDApp, two were satisfied, and one neutral (**Q7**).

The open question (**Q8**) collected the free comments that were generally positive. One of the clinicians favorably judged the reporting of the BDI-II results and the emotional analysis. He proposed to acquire other patients' data with sensors. One of them appreciated the use of correlation that is very useful for interpreting the questionnaire results. In particular, she involved three patients in this study. For one of them, inversely correlated data highlighted a real problem. For the other two, the correlation was moderately positive for both speech and image, according to the questionnaire scores. Another clinician involved two patients: in one case, the correlation was very weak positive, i.e., this result did not provide useful indications; the other patient got a moderate positive correlation for speech and strongly positive for audio, and also his judgement agreed with the questionnaire score. He also expressed a favorable judgment on the pleasantness of the EDApp interface.

EDApp got as NPS score 50% (see Fig. 9), e.g., five promoters and one detractor, who scored 5.

4.4.2 RQ2: Is the User experience adequate from the patient point-of-view?

The statistics related to UEQ questionnaire filled in by the patients are reported in Table 11. The score range is [-3,3], so we can say that all the scales have an average greater than 1 and also all the standard deviation are less than 1.2.

Table 12: Results of the UEQ Benchmark.

Scale	Comparison	Interpretation
Attractiveness	Good	10% of results better, 75% of results worse
Perspicuity	Good	10% of results better, 75% of results worse
Efficiency	Excellent	In the range of the 10% best results
Dependability	Above Average	25% of results better, 50% of results worse
Stimulation	Above Average	25% of results better, 50% of results worse
Novelty	Excellent	In the range of the 10% best results

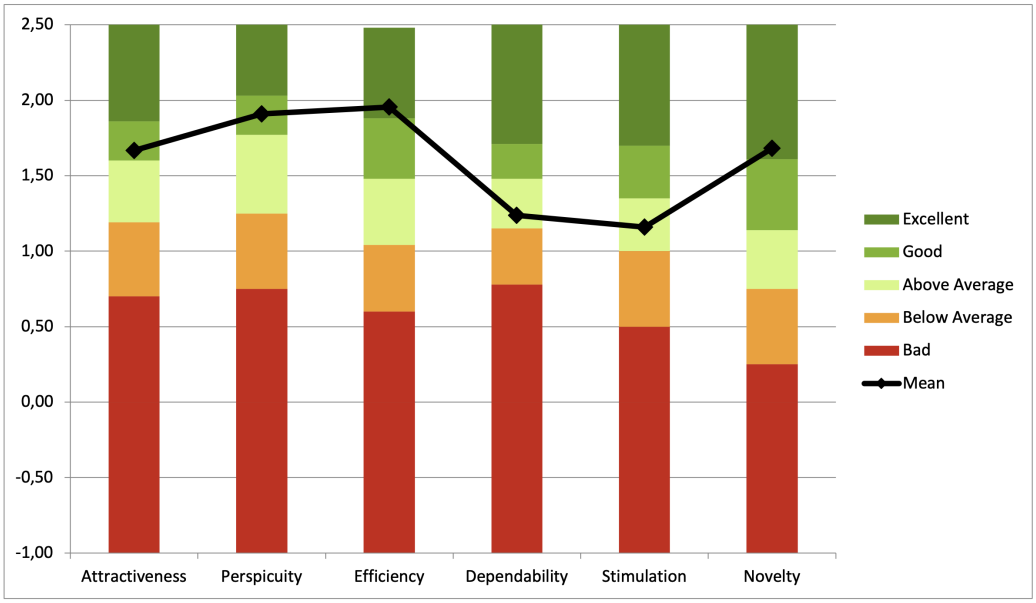


Fig. 10: A screenshot of the results of the comparison of the UEQ questionnaire with the UEQ benchmark.

The results obtained by inserting the questionnaire data in the tool proposed in [35] are shown in Table 12 and Fig. 10.

The UEQ tool puts in relation the means of the collected data for each scale with values from a benchmark data set, containing data from 20,190 people from 452 studies concerning different products (business software, web pages, webshops, social networks). In this way, we can compare EDApp scores with the scores of other products. By analyzing Figure 10 we can see that the User Experience of EDApp when compared with one of the other products reaches results above the average for Stimulation and Dependability, a good result

for Attractiveness and Perspicuity (only 10% of the other products performs better) and is excellent from Efficiency and Novelty (it is in 10% of the best results),

4.5 Threats to validity

We discuss the threats that could have affected the validity of the results in the user study by following the guidelines of Wohlin et al. [44]. In particular, we identified the following threats.

Internal Validity. A possible threat is a voluntary participation in the study (*selection threat).* But clinicians were motivated in the search of a supporting tool for depression screening and patients were in therapy to solve their problems, so they should also be motivated in filling in the questionnaires.

Construct Validity. To mitigate *evaluation apprehension threat*, we reassured participants that their data were treated anonymously and in aggregate form. It is worth mentioning that we asked the participants to sign in a consent form to use their data in anonymous form.

Conclusion Validity. To mitigate the *Reliability of measures* we used well-known and widely used measures and standard questionnaires and the NPS score. We also adopted the UEQ benchmark for comparing the obtained results w.r.t. a wider sample of tools. This benchmark has been adopted in previous work on case of applications for mental health, see for example [7]. It is important to point out that this benchmark includes studies concerning very heterogeneous tools and this may be a threat when considering evaluation provided for tools of a different category.

External Validity. The number of clinicians and patients who participated might affect this kind of validity, e.g., their cultural background and their technical skills maybe not representative of the wider population.

5 Discussion and Limitations

The preliminary evaluation results revealed that both the clinicians and the patients had a good impression of the tool: out of eight clinicians, five were promoters (see Table 9); that is, they actively would recommend the tool to their colleagues. A 50% NPS score is an excellent result if we consider that a positive score denotes that the product has more promoters than detractors. It is worth mentioning that, referring to the NPS score of big companies, Netflix had an NPS of 64, PayPal scored 63, Amazon 54, Google 53, and Apple 49[11]. Obviously, these are big companies. It should be more interesting to compare the NPS score of EDApp with systems more similar to it, but NPS is not computed in the related works. Moreover, seven clinicians judged the system as helpful and would use it in their daily activities. The patients also positively

[11] https://www.hotjar.com/net-promoter-score/

judged the tool, specifically the Efficiency and the Novelty, probably because of the use of emotion detection techniques.

A limitation of the tool may be due to the use of smartphone images for detecting emotions, while the CNN accuracy was tested on a the FER2013 dataset, where the face is carefully framed. Mobile phones detect faces in a different way, they can be too close or too far or can take only a part of the face. Before conducting the study we were not able to find a dataset with phone picture labeled with emotions, but an appropriate dataset may be created for improving the tool.

Another improvement could be the use of sensors for further supporting emotion detection. The information they collect, such as temperature, humidity, heart-Beat rate, could be integrated with images and audio. This study was conducted in the pandemic period. To collect sensor data with sensors by using our tool presented in [13] would have required us to be in presence, because it is only a prototype, not easy to use. For this reason, we preferred to use pictures and speech in this preliminary version, reassured that speech is largely adopted in recent study on depression, see for example ([29][36][23][43]).

Another limitation may be due to the the accuracy we got on the Demos dataset (71.8%). A higher accuracy has been reached by training the system of different datasets, as in [10], where the language is English and the accuracy is between 80.6% and 84.5%. Thus there may be space for improving our result.

The restricted use of the data, which are only available on the clinician device also limited the analysis. Future work may be addressed to acquire the data in collaboration with a health provider to contribute to the patient health record. Also, the acquisition of the patient data may be adopted for creating a dataset for depressed Italian people.

It is also important to point out that EDApp detects sadness. Even if sadness is considered an indicator of depression by most authors [27], sadness and depression are not the same thing, as represented in Fig. 1. In any case, EDApp does no provide any therapeutic suggestions: it only alerts the clinician in performing a deeper control in case of discordant results.

6 Conclusion and future work

When filling in a questionnaire for depression screening, the respondent may exaggerate or minimize the answers. Based on this fact, in this paper, we presented EDApp, a mobile application for supporting the clinician in the depression screening by assessing the relationship between the patient's emotion and the Beck Depression Inventory-II (BDI-II) scores. Emotions are collected by analyzing speech and images of the patient, which are processed by Deep learning techniques based on Convolutional Neural Networks. We conducted a preliminary evaluation that indicates a favorable attitude of clinicians towards the acceptability of EDApp in clinical use, providing a good Net Promoter Score (50%). The patients' user experience was positive too.

In this preliminary evaluation, we involved a reduced number of participants (also because of Covid-19 pandemic) and adopted an existing dataset for evaluating the accuracy of the neural networks. In the future, we plan to conduct testing in the wild of the system with a wide sample of participants, requiring accurate data management procedures to handle the patients' images while respecting privacy constraints. The emotion detection accuracy, especially for the speech case, should also be enhanced. This may be obtained by experimenting with other deep learning approaches or by creating larger datasets. In particular, in the image case, performance may improve by considering mages taken by the mobile device. Emotion detection may also be improved by collecting physiological data with sensors, such as the heart rate, body temperature, blood pressure, and respiratory rate. Further study may also be devoted to improve the report for the clinician, by comparing and simplifying the different visualization metaphors by eliminating redundant information.

Acknowledgement

Many thanks to all the anonymous participants involved in the evaluation and to the anonymous reviewers that largely improved the paper quality with their valuable suggestions.

Funding and/or Conflicts of interests/Competing interests

No funding was received for conducting this study.

The authors have no financial or proprietary interests in any material discussed in this article.

Data availability statement

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Ethical concerns

The evaluation procedure described in this paper obtained the ethical approval n. **0001** by the Ethic Committee of the Computer Science Department of the University of Salerno.

References

1. A. M. Badshah, J. Ahmad, N. Rahim, and S. W. Baik. Speech emotion recognition from spectrograms with deep convolutional neural network. In *2017 international conference on platform technology and service (PlatCon)*, pages 1–5. IEEE, 2017.

2. A. T. Beck. *Cognitive therapy of depression*. Guilford press, 1979.
3. A. T. Beck, R. A. Steer, and G. Brown. Beck depression inventory–ii. *Psychological Assessment*, 1996.
4. A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh. An inventory for measuring depression. *Archives of general psychiatry*, 4(6):561–571, 1961.
5. D. Bertero and P. Fung. A first look into a convolutional neural network for speech emotion detection. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5115–5119. IEEE, 2017.
6. W. C. de Melo, E. Granger, and A. Hadid. Depression detection based on deep distribution learning. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4544–4548, 2019.
7. K. Denecke, S. Vaaheesan, and A. Arulnathan. A mental health chatbot for regulating emotions (sermo) - concept and usability test. *IEEE Transactions on Emerging Topics in Computing*, 9(3):1170–1182, 2021.
8. M. Deshpande and V. Rao. Depression detection using emotion artificial intelligence. In *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, pages 858–862. IEEE, 2017.
9. P. Ekman. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999.
10. M. Ezz-Eldin, A. A. M. Khalaf, H. F. A. Hamed, and A. I. Hussein. Efficient feature-aware hybrid model of deep learning architectures for speech emotion recognition. *IEEE Access*, 9:19999–20011, 2021.
11. A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. Gailey, and C. Levinton. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. *Journal of psychiatric research*, 27(3):309–319, 1993.
12. R. Francese and P. Attanasio. Supporting depression screening with multimodal emotion detection. In A. D. Angeli, L. Chittaro, R. Gennari, M. D. Marsico, A. Melonio, C. Gena, L. D. Russis, and L. D. Spano, editors, *Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter, CHItaly '21, Bozen-Bolzano, Italy, and online (www), July 11-13, 2021*, pages 7:1–7:8. ACM, 2021.
13. R. Francese, M. Risi, and G. Tortora. A user-centered approach for detecting emotions with low-cost sensors. *Multim. Tools Appl.*, 79(47):35885–35907, 2020.
14. N. Hamiditabar, A. Chalechale, and S. J. Kabudian. Determining the severity of depression in speech based on combination of acoustic-space and score-space features. In *2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, pages 1–5, 2022.
15. D. Hamilton, J. V. Lane, P. Gaston, J. Patton, D. Macdonald, A. Simpson, and C. Howie. Assessing treatment outcomes using a single question: the net promoter score. *The bone & joint journal*, 96(5):622–628, 2014.
16. J. Hauke and T. Kossowski. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, 30(2):87–93, 2011.
17. L. He and C. Cao. Automated depression analysis using convolutional neural networks from speech. *Journal of biomedical informatics*, 83:103–111, 2018.
18. A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang. Artificial intelligent system for automatic depression level analysis through visual and vocal expressions. *IEEE Transactions on Cognitive and Developmental Systems*, 10(3):668–680, 2017.
19. A. Korszun. Facial pain, depression and stress–connections and directions. *Journal of oral pathology & medicine*, 31(10):615–619, 2002.
20. L. Laflamme, J. Chipps, H. Fangerau, N. Juth, F. Légaré, H. Sawe, and L. Wallis. Targeting ethical considerations tied to image-based mobile health diagnostic support specific to clinicians in low-resource settings: the brocher proposition. *Global health action*, 12(1):1666695, 2019.
21. C. Lemey, M. E. Larsen, J. Devylder, P. Courtet, R. Billot, P. Lenca, M. Walter, E. Baca-García, and S. Berrouiguet. Clinicians' concerns about mobile ecological momentary assessment tools designed for emerging psychiatric problems: Prospective acceptability assessment of the memind app. *Journal of medical Internet research*, 21(4):e10111, 2019.

22. L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco. Emothaw: A novel database for emotional state recognition from handwriting and drawing. *IEEE Transactions on Human-Machine Systems*, 47(2):273–284, 2017.
23. X. Lu, D. Shi, Y. Liu, and J. Yuan. Speech depression recognition based on attentional residual network. *Frontiers in Bioscience-Landmark*, 26(12):1746–1759, 2021.
24. Q. Mao, M. Dong, Z. Huang, and Y. Zhan. Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE transactions on multimedia*, 16(8):2203–2213, 2014.
25. A. McPherson and C. Martin. A narrative review of the beck depression inventory (bdi) and implications for its use in an alcohol-dependent population. *Journal of Psychiatric and Mental Health Nursing*, 17(1):19–30, 2010.
26. R. R. Morris, S. M. Schueller, and R. W. Picard. Efficacy of a web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. *Journal of medical Internet research*, 17(3):e4167, 2015.
27. S. Mouchet-Mages and F. J. Baylé. Sadness as an integral part of depression. *Dialogues in clinical neuroscience*, 2022.
28. M. M. Mukaka. A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, 24(3):69–71, 2012.
29. A. Mulay, A. Dhekne, R. Wani, S. Kadam, P. Deshpande, and P. Deshpande. Automatic depression level detection through visual input. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pages 19–22. IEEE, 2020.
30. R. Nakamura and Y. Mitsukura. Feature analysis of electroencephalography in patients with depression. In *2018 IEEE Life Sciences Conference (LSC)*, pages 53–56. IEEE, 2018.
31. M. Niu, J. Tao, B. Liu, J. Huang, and Z. Lian. Multimodal spatiotemporal representation for automatic depression level detection. *IEEE Transactions on Affective Computing*, 2020.
32. A. Pampouchidou, O. Simantiraki, C.-M. Vazakopoulou, C. Chatzaki, M. Pediaditis, A. Maridaki, K. Marias, P. Simos, F. Yang, F. Meriaudeau, et al. Facial geometry and speech analysis for depression detection. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1433–1436. IEEE, 2017.
33. E. Parada-Cabaleiro, G. Costantini, A. Batliner, M. Schmitt, and B. W. Schuller. Demos: An italian emotional speech corpus. *Language Resources and Evaluation*, pages 1–43, 2019.
34. J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
35. M. Schrepp. User experience questionnaire handbook. *All you need to know to apply the UEQ successfully in your project*, 2015.
36. D. Shi, X. Lu, Y. Liu, J. Yuan, T. Pan, and Y. Li. Research on depression recognition using machine learning from speech. In *2021 International Conference on Asian Language Processing (IALP)*, pages 52–56, 2021.
37. B. Sumali, Y. Mitsukura, Y. Tazawa, and T. Kishimoto. Facial landmark activity features for depression screening. In *2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pages 1376–1381, 2019.
38. M. Tadalagi and A. M. Joshi. Autodep: automatic depression detection using facial expressions based on linear binary pattern descriptor. *Medical & Biological Engineering & Computing*, pages 1–16, 2021.
39. J. D. Tariman, D. L. Berry, B. Halpenny, S. Wolpin, and K. Schepp. Validation and testing of the acceptability e-scale for web-based patient-reported outcomes in cancer care. *Applied Nursing Research*, 24(1):53–58, 2011.
40. M. Tasnim and E. Stroulia. Detecting depression from voice. In *Canadian Conference on Artificial Intelligence*, pages 472–478. Springer, 2019.
41. J. Torous, J. Onnela, and M. Keshavan. New dimensions and new tools to realize the potential of rdoc: digital phenotyping via smartphones and connected devices. *Translational psychiatry*, 7(3):e1053–e1053, 2017.

42. M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic. Avec 2014: 3d dimensional affect and depression recognition challenge. In *Proceedings of the 4th international workshop on audio/visual emotion challenge*, pages 3–10, 2014.
43. L. Verde, G. Raimo, F. Vitale, B. Carbonaro, G. Cordasco, S. Marrone, and A. Esposito. A lightweight machine learning approach to detect depression from speech analysis. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 330–335, 2021.
44. C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.
45. L. Yang. Multi-modal depression detection and estimation. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 26–30. IEEE, 2019.
46. X. Zhou, K. Jin, Y. Shang, and G. Guo. Visually interpretable representation learning for depression recognition from facial images. *IEEE Transactions on Affective Computing*, 11(3):542–552, 2018.