An Augmented Teleconsultation Platform for Depressive Disorders

DIOGO RAMALHO1,2, PEDRO CONSTANTINO1,2, HUGO PLÁCIDO DA SILVA2,3, (Senior Member, IEEE), MIGUEL CONSTATE4, AND JOÃO SANCHES1,2, (Senior Member, IEEE)

1Institute for Systems and Robotics (ISR-Lisboa), LARSyS, Instituto Superior Tecnico, University of Lisbon, 1049-001 Lisbon, Portugal
2Department of Bioengineering, Instituto Superior Tecnico, University of Lisbon, 1049-001 Lisbon, Portugal
3IT—Instituto de Telecomunicações, University of Lisbon, 1049-001 Lisbon, Portugal
4Department of Psychiatry, Hospital Beatriz Ângelo, 2674-514 Loures, Portugal

Corresponding author: João Sanches (jmrs@tecnico.ulisboa.pt)

This work was supported by Portuguese funds through FCT (Fundação para a Ciência e Tecnologia) under the projects reference UIDP/50009/2020, UIDB/50008/2020 and LARSyS_LA/P/0083/2020. It was also supported by FCT/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through national funds and when applicable co-funded European Union (EU) funds under Project DSAIPA/AI/0122/2020 AI-based Mobile Applications for Public Health Response (AIMHealth).

ABSTRACT Depressive disorders are a leading cause of morbidity and disability in the working population worldwide. Psychiatric assessment and treatment of depressive disorders may be improved by using objective behavioral and physiological biomarkers. In this work we present a psychiatry telemedicine platform where two important biomarkers of depression, eye-blinking rate and heart rate variability, are computed from the video stream and shown to the clinician in real time during teleconsultation. In order to validate our video processing methods, we use a public annotated dataset with image and photoplethysmography information, complemented by us with eye blinking information. For the eye-blinking detection, we obtained 94%±16% precision in the best scenario. For the estimation of %LF and LF/HF, the mean errors obtained are 16.9NU and 1.1, respectively, outperforming the state of the art. These results show the potential of the proposed telemedicine platform.

INDEX TERMS Augmented reality, depression, eye-blinking rate, heart rate variability, mental health, psychiatry, telemedicine.

I. INTRODUCTION

Depressive disorders are one of the leading causes of disability in society, with an estimated 322 million people affected [1]. The COVID-19 pandemic highlighted the impact of increased social isolation and restricted access to care, leading to higher rates of depression symptoms in the general population [2]. Depressive disorders are diagnosed based on a psychiatric history and mental state examination by a trained doctor. Notwithstanding, biomarkers could aid this approach by providing objective measurements, dissecting heterogeneity within the same diagnosis, helping predict response to treatment and developing targets for new interventions [3]. As telepsychiatry expands in the post COVID era [4], new digital based tools could help implement systematic, non-intrusive, automatic measurement and recording of depression biomarkers in routine clinical practice.

Psychomotor retardation is common in people with depressive disorders [5], [6], thus being one of the nine symptoms that are part of the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5) criteria for Major Depressive Disorder (MDD) diagnosis [7]. It can be manifested as slowed speech, dysfunctional cognition and decreased movement, affecting hands, legs, torso and head [8]. This psychomotor retardation has also been observed to influence blinking [9].

Heart Rate (HR) is defined as the number of heart beats per minute and Heart Rate Variability (HRV) concerns the fluctuation in the time intervals between adjacent heartbeats [10]. HR variation ensures optimal adaption to...
environmental challenges [11] and is a reflection of the many physiological factors modulating the normal rhythm of the heart, namely the coordination of autonomic, respiratory, circulatory, endocrine and mechanical influences over time [12]. It has been reported that a lower HRV may be a predictor of depressive disorders [13], [14], and also that abnormal HRV patterns may be in general associated with mental disorders [15], social cognition [16], executive function [17] and emotional regulation [18]. A reason for HRV to became popular in psychology is that it can be obtained from low-cost, non-invasive, and accessible equipment. Furthermore, in 2008, remote Photoplethysmography (rPPG) research officially started with reports that a subject’s face contains a signal sufficiently rich to measure HR under ambient light, using only a digital camera and signal processing [19]. This simple setup makes it possible to estimate HRV from a tele-consultation webcam video.

By recognizing that blinking and HRV metrics are related to depression, that they can be extracted from teleconsultation video and that the number of teleconsultations is growing, we extend the state-of-the-art by proposing a novel augmented teleconsultation platform to support assessment of depressive disorders. It contains the canonical functionality of any teleconference application, with additional functionalities especially designed for psychiatric consultation of depressive disorders. The latter consists of detection and quantification in real time of the eye-blinking rate, the HR and the HRV of the patient during the session, using only a video camera as a data source. These metrics are obtained in each session, stored by the platform and may be used by the clinician to observe the evolution of the patient. In order to validate the method used to extract these metrics, we assessed its performance using the Univ. Bourgogne Franche-Comté Remote PhotoPlethysmoGraphy (UBFC-RPPG) public dataset.

The remainder of this work is organized as follows: Section II describes related work, Section III details the system architecture of the proposed platform, Section IV describes the methods used to extract the biomarkers, Section V presents and discusses our results, Section VI provides an outline of the main findings and suggests next steps.

II. RELATED WORK

Computer-aided diagnosis (CAD) is the use of a computer-generated output to assist clinicians in the decision-making process. The CAD work concerning objective metrics to support psychiatric assessment has been developing models to classify subjects into depression groups and regression models to estimate depression scores [20], [21]. Biomarkers that may support the diagnosis come from different data sources such as behavioural smartphone data [22], social media text [23], physiological signals [24], speech [25] and video of the subject. Regarding visual cues, it is common to use features from the subject’s face such as facial expression, head movement, and characteristics from eyes and mouth [26]. For instance, Zhou et al. [27] developed an integrated multimodal system that uses webcam video, social media content and keyboard/mouse user interaction to classify depressed patients into two severity levels. However, they only used 5 depressed subjects out of 10, which is a small sample size.

Alghowinem et al. [9] found that in depressed subjects, the average distance between the eyelids was significantly smaller and the average duration of blinks was significantly longer. Al-gawwam et al. [28] proposed a decision support system to classify subjects into a depressed or non-depressed state based on eye blink features such as blinks per minute, blink duration and blink amplitude. In the dataset where the subjects’ read a passage, the classifier obtained a 92% accuracy. This suggests that eye blinking features can be used as metrics to support clinical assessment of depressive disorders.

Studies that investigate the relationship between HRV and depression typically use time-domain metrics and frequency-domain metrics. These metrics are derived from the normalized to-normal (NN) intervals, that represent the normalized time between two heartbeat detections. The most used time-domain metrics are the HR, the standard deviation of the NN intervals (SDNN) and the root mean square of the successive differences (RMSSD) of the NN intervals. Frequency-domain methods estimate the spectral power of bands of interest from the NN interval signal and three spectral components are distinguished in short-term recordings: very low frequency (VLF) in a range of 0.0033-0.04Hz, low frequency (LF) in a range of 0.04-0.15Hz and high frequency (HF) in a range of 0.15-0.40Hz. LH and HF bands are known to reflect changes in autonomic modulation of the heart period and can be represented in normalized units to emphasize the balanced behavior of the two branches of the autonomic system [29]. An LF/HF ratio can be computed from the LF and HF spectral powers, with the underlying assumption being that LF power is mainly generated by the sympathetic nervous system and HF power is mainly produced by the parasympathetic nervous system. Thus, a low LF/HF ratio can be observed in tending-befriend behaviors (parasympathetic dominance) and a high LF/HF ratio occurs in fight-or-flight situations (sympathetic dominance) [30].

Recently, comprehensive analyses of effect sizes for seven distinct resting-state HRV measures [31] were made, namely HF, LF, VLF, LF/HF, RMSSD, SDNN and mean NN interval duration. Results suggest that patients with depressive disorders are likely to display small reductions in HF, LF, VLF, RMSSD (largest effect size), SDNN and NN intervals and an increase in LF/HF ratio. They conclude that depressive states are not associated with alteration in specific indicators of HRV, but rather with abnormalities in several time- and frequency-domain measures. In [32], authors found correlation between 15-min ECG-based HRV measures and clinical state of depressive disorders. Within 62 depressed patients and 65 non-depressed controls and testing several HRV measures, they have seen a decreased HF%, amongs
others, to be highly correlated with a diagnostic of depressive disorders.

To the best of our knowledge, this work is the first to propose a real-time system that displays depression-related biomarkers to the clinician using only the subject’s teleconsultation video as input. Furthermore, this is the first work to combine eye-blinking rate with HRV for assessment of depression, in contrast to other works which either use these features alone or with other depression-related biomarkers.

III. SYSTEM ARCHITECTURE
The system is composed of two main parts: the web application and the video processing. In the web application, the clinician is able to start a teleconsultation and then invite the patient to join. After the patient joins the teleconsultation, the clinician runs the video processing script. This script will do the analysis of the biomarkers and store the results in the database in background during the teleconsultation. Finally, the clinician is able to select in the web application the results of the biomarkers’ analysis he wishes to visualize. A general overview of the system architecture can be seen in Fig. 1a.

A. WEB APPLICATION
The web application includes authentication, video meeting integration and charts for data visualization. The technological stack we used to develop the web application was Bootstrap, JavaScript and Firebase. We used the Bootstrap framework to develop the graphical layout and the SB Admin 2 Bootstrap template was used as a starting layout for the web application. Furthermore, the authentication was done using Firebase Authentication with an email address/password sign-in method. The web application was deployed using Firebase Hosting.

We chose Cloud Firestore from Firebase as our application’s database, which was used to store the results of the biomarker analysis. There is a collection called users that stores the patient’s identification string and username, which follows the anonymous pattern PSLTEC.APP1 (for patient 1). The association between username and the real name is kept in an Excel file stored in the clinician’s computer. There is another collection called consultations that stores data about all the consultations performed in the platform. For each consultation we store the identification string of the patient and the clinician, the date of the consultation and the results obtained from the biomarker analysis. This data allows retrospective comparisons between consultations and to examine the overall longitudinal course of the patient’s depression.

In order to allow the creation of a teleconsultation inside the web app we used the Zoom Web SDK. First, we created a Zoom developer account in Zoom Marketplace and then registered it as a JSON Web Tokens (JWT) App. After the registration, we obtained the API key and API secret which allow the authentication in the Zoom server when the request to start a teleconsultation is made. Then, we integrated the Zoom Web SDK by including the dependencies on the teleconsultation HTML file. Finally, we built a structure for the process of meeting creation with three components: 1) browser client; 2) proxy server; 3) Zoom server. The proxy server is a Firebase Cloud Function that exists because the API key and secret should not be stored in the browser client for security reasons. The process consists on HTTP messages exchanged between the components with the requests to start the meeting and the responses with the necessary information to do so (Fig. 1b). These messages carry the necessary authentication tokens, to allow the communication with Firebase and with Zoom. The settings of the meeting are defined in the proxy server.

We added the charts using D3.js. The eye-blinking rate, the Standard Deviation of NN Intervals (SDNN) and the HR are represented by line charts in the web application, while the LF/HF ratio is represented by a doughnut chart. In the line chart representation, the vertical axis represents the magnitude of the biomarker, such as the number of blinks. The horizontal axis represents the time instant in which the biomarker was obtained, with the axis ticks being displayed every minute. In addition to the line, every data point is also displayed by a circle with a radius larger than the line’s width. If the mouse is hovered over this circle, a tooltip shows the value and the instant associated to that circle in a blue box. An example of the charts’ visualization can be seen in Fig. 1c.

B. VIDEO PROCESSING
The video processing consists of three main steps: 1) capture the frames from teleconsultation video; 2) perform the biomarker analysis; and 3) send results of the analysis to the database.

The processing is done by a Python script running in the background on the computer, which must be started by the clinician. In the main loop, the process goes from frame capture (Step 1) to biomarker analysis (Step 2) sequentially and uninterruptedly. The frame is obtained by capturing a portion of the teleconsultation video screen (in this case the Python library MSS was used). Every time a frame is captured the analysis of each biomarker is done on this frame. Finally, the results of the biomarkers analysis are sent (Step 3) every 15 seconds to the database in parallel, and thus the main loop is not affected.

IV. BIOMARKER ANALYSIS
The biomarker analysis is the part where we take a series of frames and extract measures that may be clinically relevant. In this work, the focus was on measures that could be extracted from the face, more specifically the eye-blinking rate, the heart rate and its variability.

For every frame, the first two steps are face and landmark points’ detection, needed to compute the biomarkers. The face is detected using the dlib frontal face detector [33]. The detector receives a frame in greyscale and returns a square that fits the face of the subject. Then, the detected face is given as input to the dlib facial landmark detector [33], which outputs 68 facial landmark points, as can be seen in Fig. 2.
These points can be indexed to identify specific areas of our face such as eyes, nose, mouth and eyebrows, which can then be used to extract biomarkers. The facial landmark detector [34] is pre-trained on the 300-W dataset [35]. This dataset covers a large variation of identity, expression, illumination conditions, pose, occlusion and face size, which is relevant during a teleconsultation due to the diversity of patients and possible variations in the consultation environment.

**A. EYE-BLINKING RATE**

The eye-blinking rate is the number of blinks per time interval, which was defined as 15 seconds in this work. In order to obtain this value, three main steps were followed: 1) signal acquisition; 2) signal filtering; and 3) blink detection.

There are six landmark points per eye, as can be seen in Fig. 3. These landmark points represent \((x,y)\) coordinates in pixels in the frame, which means that distances can be calculated between them. As our eyes open and close, the vertical distances will be varying over time, as noted by Soukupová and Čech [36]. In their work, a ratio of how open or closed the eye is, the Eye Aspect Ratio (EAR), can be measured using the following formula:

\[
EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \|p_1 - p_4\|},
\]

where \(p_1, \ldots, p_6\) are the six landmark points, as represented in Fig. 1. As the eye closes, the vertical distance represented
by the numerator in the formula, will decrease its value (Fig. 3a), thus decreasing the EAR. On the other hand, as the eye opens, the vertical distance will increase (Fig. 3b) and thus the EAR will increase as well. The EAR is calculated for each eye and then the average is obtained. If the EAR is calculated for each frame, an EAR signal is generated, as can be seen in Fig. 4.

The EAR signal is noisy and needs to be filtered. We chose a low-pass filter as it is fast and inexpensive, which is necessary in real-time applications. The eye blinking duration has been characterized to last $572 \pm 25$ ms [38] and thus the cutoff frequency was heuristically set to 3Hz.

To identify the blinks we use a detector with an adaptive threshold. Every time the EAR is smaller than the threshold, the detector predicts a new blink. The average open EAR and the EAR when a subject blinks varies between individuals and thus a threshold that adjusts automatically given the past signal characteristics is ideal for this detection. The detection threshold is updated given the previous maximum and minimum EAR values. More specifically,

$$d_{thr}[n] = EAR_{min}[n] + r_{thr}(EAR_{max}[n] - EAR_{min}[n])$$  (2)

where $d_{thr}[n]$ is the detection threshold at frame $n$, $EAR_{max}[n]$ the maximum EAR in the filtered signal until frame $n$, $EAR_{min}[n]$ the minimum in the filtered signal until frame $n$ and $r_{thr}$ is a parameter that controls how much of the difference will be summed to $EAR_{min}[n]$. In Fig. 5, the threshold can be observed in blue, as a sum to the red line of 20% of the difference between the green and red lines.

The detection threshold requires calibration with the first blink and thus its initialization is made assuming that the difference between $EAR_{max}[n]$ and $EAR_{min}[n]$ is higher than a certain value that we called starting window ($s_w$). The $r_{thr}$ and the $s_w$ parameters were chosen experimentally as will be explained further.

### B. HEART RATE AND HEART RATE VARIABILITY

The extraction of real-time HR and HRV from video is accomplished with seven steps: 1) Region of Interest (ROI) selection; 2) RGB signal extraction; 3) Resampling; 4) rPPG signal extraction, 5) Filtering, 6) Peak detection and 7) HR and HRV estimates.

The ROI selection is based on facial landmarks and allows to segment skin regions with an accessible pulse signal. Here, the regions comprising the cheeks, the nose and the forehead are considered, while the regions for beard and eyes are removed (Fig. 6).

The starting point is the computation of the average intensities of each color component within the ROI for each frame $n$,

$$a_n = (r_n, g_n, b_n)$$  (3)

Because the Zoom session does not guarantee an uniform frame rate, a resampling step is introduced. To obtain a uniformly sampled color average linear interpolation is used.

To extract the one-dimensional rPPG signal from three-dimensional RGB signal the POS method is used [39]. This is a chrominance-based method, which means it uses skin tone knowledge a priori, thus requiring less knowledge of the rPPG signature while staying tolerant to distortion. The POS method is a sliding window algorithm. Here, window size, $POS = f_s \cdot POS_s$, was chosen to correspond to $POS_s = 1.6$ seconds, which encapsulates one cardiac cycle as recommended by the authors. This choice introduces only a 1.6 seconds delay from acquiring the frame to extracting the corresponding facial blood volume, which is short enough for real-time applications.

The next step is filtering the raw rPPG signal. Filtering is performed through a second-order Infinite Impulse Response (IIR) bandpass butterworth filter, $[0.8, 2.5]$ Hz. These cut-off
Several frequency-domain metrics are considered: normalized LF power, normalized HF power and LF/HF ratio. The Lomb-Scargle method [42] is used for deriving frequency-domain estimations. This method is less computationally expensive than others [43], which suits the real-time scenario.

V. BIOMARKER VALIDATION

The proposed methods were validated in the UBFC-RPPG video dataset [37]. The dataset includes 40 videos with a frame rate of 30 frames-per-second and a resolution of 640 × 480. In the videos, the subject sits in front of the camera (about 1 meter away from the camera) with his/her face visible for more than one minute, which is a similar setting to a teleconsultation. The dataset includes ground truth signals acquired using finger PPG, running parallel to the video capture. Even though the dataset was created for the rPPG field, it is also a good fit for blinking research. In this sense, we manually labeled the dataset, providing for each video a text file containing the frames matching blinking events. We make this ground truth available at Zenodo [44].

A. EYE-BLINKING

In order to evaluate the performance of the blink detection model, we compared the blinking ground truth of the dataset with our model predictions. We obtained the ground truth by manually retrieving the number of blinks and the frame associated to each blink. We considered the frames where the eyes were more closed, which are coincident with the negative peak of the signal. Then, our blinking detection method created a list with the predictions of the blinking events for each video. In order to check whether the ground truth and prediction events are coincident, a five-frame error temporal window was considered, because it is considered the time it takes for a blink.

When comparing both, three situations were considered: true positive, false negative and false positive. A true positive occurs when the model predicts a blink and it occurred. A false positive occurs when the model predicts a blink and it did not exist. A false negative occurs when the model did not predict a blink and it existed. The true negatives were ignored as they occur when the model did not predict a blink and it did not exist, which would give many events. In Fig. 8 we can see two true positive events in Subject 1, with the predictions represented as red peaks and the ground truth represented as green peaks. The metrics used to evaluate the performance are precision, recall and F1-score, as they are not affected by the imbalance in events.
The performance of the proposed detector was assessed with five different $r_{thr}$ and seven different $s_w$. We concluded that the best detector uses $s_w = 0.07$ and $r_{thr} = 0.2$, as it was the one presenting higher F1-score. We only took in account the predicted events after the first true blink, ignoring the false positives that occur before the detection threshold calibration.

The quality of the facial landmark prediction in the dataset decreased with subjects that use glasses, have hair occluding the eyes and half-close the eyes for long periods of time. Considering this, the model’s performance was assessed in the whole dataset (Dataset A1 - 39 subjects) and in two subsets of this dataset. One that excludes subjects with glasses (Dataset A2 - 29 subjects) and a second one that excludes subjects that satisfy all the previous conditions (Dataset A3 - 25 subjects). The average (AVG) precision was 75% in Dataset A1, 84% in Dataset A2 and 94% in Dataset A3. These results can be seen in Table 1.

After observing the precision results of each subject individually, it was observed that the standard deviation (SD) was very large (Table 1). In order to better understand the standard deviation, the precision results of each individual for the three datasets were sorted in descending order and plotted in Fig. 9. In this plot, it can be seen that the percentage of subjects where the model had 100% of precision is approximately 60% in dataset A1, 70% in dataset A2 and 80% in dataset A3. This shows that the proposed detector performs very well in subjects under normal conditions. The obtained results were not compared to other works because we’re the first to validate a blinking method in the UBFC-RPPG dataset.

### B. HEART RATE AND HEART RATE VARIABILITY

The performance of the proposed rPPG method is reported in terms of mean absolute error between estimated and ground truth HR and HRV features. For each video of the dataset, we computed the average HRV metrics for the full record and the HR for 15-second windows, 30-second windows and the full record. This was done for both the face rPPG signal (estimations) and the finger PPG signal (ground truth).

Then, we computed the mean absolute error between the HR estimations and the ground truth HR values (Table 2).

Comparing against the state-of-the-art in HR estimation (Table 3), the proposed method ($\Delta HR = 2.16$ bpm) performs extremely well, competing with the best previous non-supervised or supervised methods. Regarding HRV estimations, the method performs well when compared to state-of-the-art reports (see Table 3). For the SDNN and RMSSD, the method performs in line with the state-of-the-art, in particular [41] which is the best performing method in the literature. For %LF (NU) and LF/HF estimation, the proposed method outperforms all methods reported.

According to [30], average short-term human HRV features have the following ranges: SDNN ranges from 32-93 ms; RMSSD ranges from 19-75 ms; %LF (NU) ranges from 30-65 and LF/HF ranges from 1.1-11.6. For SDNN, RMSSD and %LF the proposed method performs with an average error of 20.5 ms, 39.2 ms and 16.9 NU, respectively. This suggests that the method should be able to accomplish only a qualitative assessment of these features. For LF / HF ratio, the mean error is the smallest compared to its respective
human range, making it the most reliable HRV biomarker coming out of the proposed method.

VI. CONCLUSION
The psychiatric approach on depressive disorders is based on an interaction between patient and clinician that uses subjective measures to diagnose and follow-up. Biomarkers can aid this approach by providing objective measurements. This paper presents a telemedicine platform that provides depression biomarkers in real-time during a teleconsultation. This platform allows the clinician to start a teleconsultation with a patient using the teleconference platform Zoom and visualize in almost real time results of blinking and heart rate variability analysis. The results are shown in charts and the clinician is able to select the results he wishes to visualize. The results of the analysis are stored in a database and can be analysed after the teleconsultation ends.

The video processing methods were validated on the UBFC-RPPG dataset. The videos in the UBFC-RPPG dataset resemble a teleconsultation setting, reason for which we chose this dataset to develop our method. HRV estimation from webcam video looks promising, and the proposed real time method performs very well compared to the state of the art. Hence, rPPG can become an important tool for teleconsultation and depressive disorders research. Future work will aim to test our psychiatry telemedicine platform and validate the incorporated biomarker measurements in patients diagnosed with depression. Emphasis will also be given to studying the effect of head movement on the proposed distance-based (eye aspect ratio) and luminosity-based metrics (HR and HRV).

REFERENCES
[1] M. J. Friedrich, “Depression is the leading cause of disability around the world,” JAMA, vol. 317, no. 15, p. 1517, Apr. 2017.
[2] A. Dawel, Y. Shou, M. Smithson, N. Cherbuin, M. Banfield, A. L. Calear, L. M. Farrer, D. Gray, A. Gulliver, T. Houser, S. M. McCullum, A. R. Morse, K. Murray, E. Newman, R. M. R. Harris, and P. J. Batterham, “The effect of COVID-19 on mental health and wellbeing in a representative sample of Australian adults,” Frontiers Psychiatry, vol. 11, Oct. 2020, Art. no. 579985.
[3] A. M. Buch and C. Liston, “Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics,” Neuropsychopharmacology, vol. 46, no. 1, pp. 156–175, Jan. 2021.
[4] H. Li, A. Gleave, A. Kent-Wilkinson, D. Leidl, M. Kleib, and T. Rishing, “Transition of mental health service delivery to telepsychiatry in response to COVID-19: A literature review,” Psychiatric Quarterly, vol. 93, no. 1, pp. 181–197, Jun. 2021.
[5] J. E. Greisen and B. J. Carroll, “Psychomotor function in affective disorders: An overview of new monitoring techniques,” Am. J. Psychiatry, vol. 138, no. 11, pp. 1441–1448, Nov. 1981.
[6] D. J. Widlocher, “Psychomotor retardation: Clinical, theoretical, and psychometric aspects,” Psychiatric Clinics North Amer., vol. 6, no. 1, pp. 27–40, Mar. 1983.
[7] Diagnostic and Statistical Manual of Mental Disorders: DSM-5™, American Psychiatric Publishing, Washington, DC, USA, 2017.
[8] J. S. Buyukdura, S. M. McClintock, and P. E. Croarkin, “Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment,” Prog. Neuro-Psychopharmacol. Biol. Psychiatry, vol. 35, no. 2, pp. 395–409, Mar. 2011.
[9] S. Alghowinem, R. Goecke, M. Wagner, G. Parker, and M. Breakspear, “Eye movement analysis for depression detection,” in Proc. IEEE Int. Conf. Image Process., Sep. 2015, pp. 4220–4224.
[10] R. McCraty and F. Shaffer, “Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk,” Global Adv. Health Med., vol. 4, no. 1, pp. 46–61, Jan. 2015.
[11] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, “Heart rate variability: A review,” Med. Biol. Eng. Comput., vol. 44, no. 12, pp. 1051–1051, Dec. 2006.
[12] D. S. Quintana, G. A. Alves, and J. A. J. Heathers, “Guidelines for reporting articles on psychiatry and heart rate variability (GRAPH): Recommendations to advance research communication,” Transl. Psychiatry, vol. 6, no. 5, p. e803, May 2016.
[13] A. R. Brunoni, A. H. Kemp, E. M. Dantas, A. C. Goulart, M. A. Nunes, P. S. Boggio, J. G. Mill, P. A. Lotufo, F. Fregni, and I. M. Bensenhor, “Heart rate variability is a trait marker of major depressive disorder: Evidence from the sertraline vs. electric current therapy to treat depression clinical study,” Int. J. Neuropsychopharmacol., vol. 16, no. 9, pp. 1937–1949, Oct. 2013.
[14] I. H. Ha, S. Park, D. Yoon, and B. Kim, “Short-term heart rate variability in older patients with newly diagnosed depression,” Psychiatry Res., vol. 226, nos. 2–3, pp. 484–488, Apr. 2015.
[15] G. A. Alves, D. S. Quintana, I. B. Hickie, and A. J. Guastella, “Autonomic nervous system dysfunction in psychiatric disorders and the impact of psychotropic medications: A systematic review and meta-analysis,” J. Psychiatry Neurosci., vol. 41, no. 2, pp. 89–104, Mar. 2016.
[16] D. S. Quintana, A. J. Guastella, T. Outhred, I. B. Hickie, and A. H. Kemp, “Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition,” Int. J. Psychophysiol., vol. 86, no. 2, pp. 168–172, 2012.
[17] A. L. Hansen, B. H. Johnsen, J. J. Sollers, K. Stenf, and J. F. Thayer, “Heart rate variability and its relation to prefrontal cognitive function: The effects of training and detraining,” Eur. J. Appl. Physiol., vol. 93, no. 3, pp. 263–272, Dec. 2004.
[18] A. Reinecke, N. Filippini, C. Berna, D. G. Western, B. Hans, M. J. Cooper, P. Taggart, and C. J. Harmer, “Effective emotion regulation strategies improve IMRI and ECG markers of psychopathology in panic disorder: Implications for psychological treatment action,” Transl. Psychiatry, vol. 5, no. 11, p. e673, Nov. 2015.
[19] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” Opt. Exp., vol. 16, no. 26, p. 21434, Dec. 2008.
[20] M. K. Hasan, S. Khadem, and H. R. Marateb, “A comprehensive review of computer-aided diagnosis of major mental and neurological disorders and suicide: A biostatistical perspective on data mining,” Diagnostics, vol. 11, no. 3, p. 393, Feb. 2021.
[21] M. Al Jazaery and G. Guo, “Video-based depression level analysis by encoding deep spatiotemporal features,” IEEE Trans. Affect. Comput., vol. 12, no. 1, pp. 206–216, Jan. 2021.
[22] K. O. Asare, Y. Terhorst, J. Vega, E. Peltonen, E. Lagerspetz, and D. Ferreira, “Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: Exploratory study,” JMIR mHealth uHealth, vol. 9, no. 7, Jul. 2021, Art. no. e26540.
[23] R. Safa, P. Bayat, and L. Mognhchtader, “Automatic detection of depression symptoms in Twitter using multimodal analysis,” J. Supercomput., vol. 78, no. 4, pp. 4709–4744, Mar. 2022.
[24] A. Y. Kim, E. H. Jang, S. Kim, K. W. Choi, H. J. Jeon, H. Y. Yu, and S. Byun, “Automatic detection of major depressive disorder using electrodermal activity,” Sci. Rep., vol. 8, no. 1, p. 17030, Nov. 2018.
[25] L. He and C. Cao, “Automated depression analysis using convolutional neural networks from speech,” J. Biomed. Inform., vol. 83, pp. 103–111, Jul. 2018.
[26] A. Pampouchidou, P. G. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pedadiitis, and M. Tsiknakis, “Automatic assessment of depression based on visual cues: A systematic review,” IEEE Trans. Affect. Comput., vol. 10, no. 4, pp. 445–470, Oct. 2019.
[27] D. Zhou, J. Luo, V. Silenzio, Y. Zhou, J. Hu, G. Currier, and H. Kautz, “Tackling mental health by integrating unobtrusive multimodal sensing,” in Proc. 29th AAAI Conf. Artif. Intell., AAAI 27th Innov. Appl. Artif. Intell. Conf., in Proceedings of the National Conference on Artificial Intelligence, Jun. 2015, pp. 1401–1408.
[28] S. Al-gawwam and M. Benaissa, “Depression detection from eye blink features,” in Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT), Dec. 2018, pp. 388–392.
[29] T. Force, “Standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology and the north American society of pacing and electrophysiology,” Circulation, vol. 93, no. 5, pp. 1043–1065, 1996.
[30] F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” Frontiers Public Health, vol. 5, p. 258, Sep. 2017.

[31] C. Koch, M. Wilhelm, S. Salzmann, W. Rief, and F. Euteneuer, “A meta-analysis of heart rate variability in major depression,” Psychol. Med., vol. 40, no. 12, pp. 1948–1957, Sep. 2010.

[32] R. Hartmann, F. M. Schmidt, C. Sander, and U. Hegerl, “Heart rate variability as indicator of clinical state in depression,” Frontiers Psychiatry, vol. 9, p. 735, Jan. 2019.

[33] D. E. King (Dec. 2009), DLIB-ML: A Machine Learning Toolkit [Online]. Available: http://dlib.net/python/index.html

[34] V. Kazemi and J. Sullivan, “One millisecond face alignment using an ensemble of regression trees,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1867–1874.

[35] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 faces in-the-wild challenge: Database and results,” Image Vis. Comput., vol. 47, pp. 3–18, Mar. 2016.

[36] T. Soukupova and J. Cech, “Real-time eye blink detection using facial landmarks,” in Proc. Comput. Vis. Winter Workshop (CVWW), 2016, pp. 42–50.

[37] S. Bobbia, R. Macwan, Y. Benezhet, A. Mansouri, and J. Dubois, “Unsupervised skin tissue segmentation for remote photoplethysmography,” Pattern Recognit. Lett., vol. 124, pp. 82–90, Jun. 2017.

[38] K.-A. Kwon, R. J. Shipley, M. Edirisinghe, D. G. Ezra, G. Rose, S. M. Best, and R. E. Cameron, “High-speed camera characterization of voluntary eye blinking kinematics,” J. Roy. Soc. Interface, vol. 10, no. 85, Aug. 2013, Art. no. 20130227.

[39] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote PPG,” IEEE Trans. Biomed. Eng., vol. 64, no. 7, pp. 1479–1491, Jul. 2017.

[40] E. J. Argüello-Prada, “The mountaineer’s method for peak detection in photoplethysmographic signals,” Revista Facultad de Ingeniería Universidad Antioquia, vol. 90, pp. 42–50, Jan. 2019.

[41] A. Gudi, M. Bittner, and J. van Gemert, “Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation,” Appl. Sci., vol. 10, no. 23, p. 8630, Dec. 2020.

[42] S.-Y. Tseng and W.-C. Fang, “A time-frequency HRV processor using windowed Lomb periodogram,” in Database Theory and Application, Bio-Science and Bio-Technology. Berlin, Germany: Springer, 2010, pp. 273–282.

[43] G. Ramos, M. Alfaras, and H. Gamboa, “Real-time approach to HRV analysis,” in Proc. 11th Int. Joint Conf. Biomed. Eng. Syst. Technol., 2018, pp. 1–8.

[44] P. Constantino and D. Ramalho, (2022). Eye Blink Events Ground Truth for UBFC Video Dataset. Dataset Zenodo. [Online]. Available: https://zenodo.org/record/6470025#.Y5KqzHZBzIU

[45] P. Li, Y. Benezhet, K. Nakamura, R. Gomez, C. Li, and F. Yang, “An improvement for video-based heart rate variability measurement,” in Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP), Jul. 2019, pp. 433–439.

[46] R. Macwan, S. Bobbia, Y. Benezhet, J. Dubois, and A. Mansouri, “Periodic variance maximization using generalized eigenvalue decomposition applied to remote photoplethysmography estimation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2018, pp. 1332–1340.

[47] R. Song, H. Chen, J. Cheng, C. Li, Y. Liu, and X. Chen, “PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography,” IEEE J. Biomed. Health Inform., vol. 25, no. 5, pp. 1373–1384, May 2021.

PEDRO CONSTANTINO was born in Leiria, Portugal, in 1996. He received the M.S. degree in biomedical engineering from the Instituto Superior Técnico (IST-UL), Lisbon, Portugal, in 2021. His current interests include machine learning, data science, and signal processing.

HUGO PLÁCIDO DA SILVA (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the Instituto Superior Técnico (IST), University of Lisbon (UL). He has been a Researcher at the Instituto de Telecomunicações, since 2004, and has been a Co-Founder of multiple technology-based healthcare companies, since 2007. He is also a Professor at IST-UL. His current interests include biosignal research, systems engineering, signal processing, and pattern recognition. His work has been internationally distinguished with several academic and technical awards.

MIGUEL CONSTATE received the Ph.D. degree in neuropsychology of psychosis from the Institute of Psychiatry, KCL. He is currently a Consultant General Adult Psychiatrist, he completed his specialist training in the U.K., at the Maudsley Hospital, acquiring membership of the Royal College of Psychiatrists. He is also based in Lisbon, Portugal, his work is mostly devoted to providing clinical care, managing a broad range of psychiatric disorders in varied clinical settings, both in public and private sectors. Simultaneously, he is actively involved in medical teaching, psychiatry training, clinical auditing, and service development. His research interests include exploring, alongside biomedical engineers, the combined use of telepsychiatry, and mental health biomarkers in routine clinical psychiatric care.

DIOGO RAMALHO was born in Lisbon, Portugal, in 1998. He received the B.S. and M.S. degrees in computer science and engineering from the Instituto Superior Técnico (IST-UL), Lisbon, in 2019 and 2021, respectively. His research interests include systems engineering, data science, and telemedicine.

JOÃO SANCHES (Senior Member, IEEE) received the master’s degree in biomedical engineering. He is currently a Full Professor at the Department of Bioengineering (DBE), Instituto Superior Técnico, University of Lisbon, where he is coordinating the scientific area of biomedical systems and biosignals. He is also a Senior Researcher and a member of the Board of Directors of the Institute for Systems and Robotics (ISR). His work has been focused on statistical signal and image processing of biomedical data in applications for diagnosis and follow-up in cancer and mental health. He has published more than 170 international publications (ORCID and Google Scholar). He is a Senior Member of the IEEE Engineering in Medicine and Biology Society (EMBS).

VOLUME 10, 2022

130571