Indonesian Journal of Electrical Engineering and Computer Science
Vol. 19, No. 3, September 2020, pp. 1677~1684
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v19.i3.pp1677-1684

On the review of image and video-based depression detection using machine learning

Arselan Ashraf[1], Teddy Surya Gunawan[2], Bob Subhan Riza[3], Edy Victor Haryanto[4], Zuriati Janin[5]
[1,2]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia
[2,3,4]Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Indonesia
[5]Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received Jan 7, 2020
Revised Mar 10, 2020
Accepted Mar 24, 2020

Keywords:

Data acquisition
Depression database
Depression prediction
Machine learning

ABSTRACT

Machine learning has been introduced in the sphere of the medical field to enhance the accuracy, precision, and analysis of diagnostics while reducing laborious jobs. With the mounting evidence, machine learning has the capability to detect mental distress like depression. Since depression is the most prevalent mental disorder in our society at present, and almost the majority of the population suffers from this issue. Hence there is an extreme need for the depression detection models, which will provide a support system and early detection of depression. This review is based on the image and video-based depression detection model using machine learning techniques. This paper analyses the data acquisition techniques along with their databases. The indicators of depression are also reviewed in this paper. The evaluation of different researches, along with their performance parameters, is summarized. The paper concludes with remarks about the techniques used and the future scope of using the image and video-based depression prediction.

Corresponding Author:

Teddy Surya Gunawan,
Department of Electrical and Computer Engineering,
International Islamic University Malaysia, Malaysia.
Email: tsgunawan@iium.edu.my

1. INTRODUCTION

Machine learning is the technique of data analysis that encompasses a computer to learn to classify or predict the given input to produce a smart decision or result as output. Since machine learning has found its significance in almost all the prominent fields, and it has been implemented in the medical diagnostics as well to detect and classify the various diseases, including neurodegenerative diseases [1]. The inclusion of machine learning in the field of medicine has shown an increase in prediction accuracy as compared to other existing conventional techniques. Machine learning technique has also been implied to study the changes in the emotions, voice, and facial expressions.

Depression is the most common type of physiological or mood disorder affecting many individuals around the globe. Depressed people are more prone to many other problems, like sadness, loneliness, and anxiety. It is challenging for people suffering from depression to concentrate on their work, communicate with the people, and much more [2]. Hence depression detection is very vital for the proper assessment and the treatment of an individual. Machine learning incorporates many techniques to detect depression in an individual via sound or video recordings. Studies have revealed that depressed people have a difference in their physical and physiological features. These features are extracted and fed to the machine learning algorithm to determine the possible result. There should be a sufficient amount of knowledge regarding the subject to make a well-fitted model for detecting depression. Many supervised machine learning techniques can be used to make a prediction model, including Convolution Neural Networks (CNNs), Support Vector

Machine (SVM), Recurrent Neural Networks (RNNs), and Deep Neural Networks (DNNs) [3]. Experts are continuously working to build a valid model for the detection of depression by using different approaches and completely different data sources. The images/videos can be used to extract the information of the face and analyze the attributes to train a machine learning model to detect the depression when subjected to testing. There are many filters to extract facial features like the Gabor filter and then classify these features using any classification algorithms [4].

This paper aims to review the different techniques in detecting depression using image/video as the source for feature extraction. Different datasets used will be discussed under the broader category of data acquisition. The general architecture of the machine learning model using the neural networks will be briefly examined. Different image and video indicators are reviewed under the shadow of indicators of depression. Moreover, many evaluation criteria are examined to determine the performance parameters of the work. Finally, a brief conclusion is provided regarding the review along with the future work that can be done in the same field to improve the benchmarks of precision, accuracy, and time in the successive machine learning models for the depression detection.

2. IMAGE AND VIDEO DATABASE FOR DEPRESSION DETECTION

2.1. Data acquisition

The depression detection using machine learning models requires sufficient data to train the model properly. Some create their datasets using their collection of image/video samples, and others might use the image/video databases which have already been created for the researchable use. The process of collecting their own image/video data for creating their own datasets is mostly based on mental or physiological questioning and recording the response. These recordings or images are then subjected to feature extraction algorithms. The features can vary from one person to another. However, the ten general features are listed in [5], including pupil dilation/bias, iris movement, eyelid activity (openings, blinking), eye gaze (limited and shorter eye contact), mouth corners angled down, smile intensity and duration, facial activity, sad/negative/neutral expression occurrence, head pose (orientation, movement), and facial expression occurrence (variability and intensity).

The dataset created by the National Institute of Health [6] to diagnose the depression among patients have several variables as follows:

a) Hospt: The patient's hospital, shown by the code of 5 emergency clinics (1, 2, 3, 5, or 6)
b) Treat: The treatment obtained by the patient (Lithium, Imipramine, or Placebo)
c) Outcome: Whether or not a repeat happened during the patient's treatment (Recurrence or No Recurrence)
d) Time: Either the time (days) till repeat, or if no-repeat, the length (days) of the patient's cooperation in the examination.
e) AcuteT: The time (days) that the patient was discouraged preceding the investigation.
f) Age: The age of the patient in years when the patient entered the investigation.
g) Gender: The patient's gender (1 = Female, 2 = Male)

2.2. Depression databases

Depression detection is interdisciplinary, involving computer science, physiology, interaction management, and linguistics. It is not always easy to incorporate all these skill sets to create a dataset. The solution to this is to use a publically available dataset for the researchers. The list of the image/video depression databases available for the researchers is shown in Table 1.

Table 1. Publicly available datasets for depression detection

Author(s)-Year	Dataset	Modality	Depression Label Annotation
Valstar et al. (2013) [7]	AVEC 2013	Video/audio	Self-report survey (BDI-II)
Valstar et al. (2014) [8]	AVEC 2014	Video/audio	Self-report survey (BDI-II)
Lieberman & Meyer (2014) [9]	Crisis Text Line	Text	Crisis counselor judgment
Gratch et al. (2014) [10]	DAIC	Video/audio	Self-report survey (PHQ-8)
Becker et al. (1994) [11]	DementiaBank Database	Video/audio	Clinical diagnosis of depression (HAM-D)
Milne et al. (2016) [12]	ReachOut Triage Shared Task	Text	Expert judged for crisis/green/amber/red
Pradhan et al. (2014) [13]	SemEval-2014 Task 7	Text	Hand labeled for depression

3. DEPRESSION DETECTION METHODS

Detection of depression from the images/videos requires a clear and precise definition of the depressed face [14]. There should be a proper differentiation between depressed and sad features because, in both cases, some of the facial features resemble. Most of the work in this field is done on the datasets collected from adult patients since they possess the majority of this cause. The general design for the detection of depression using machine learning architecture is shown in Figure 1.

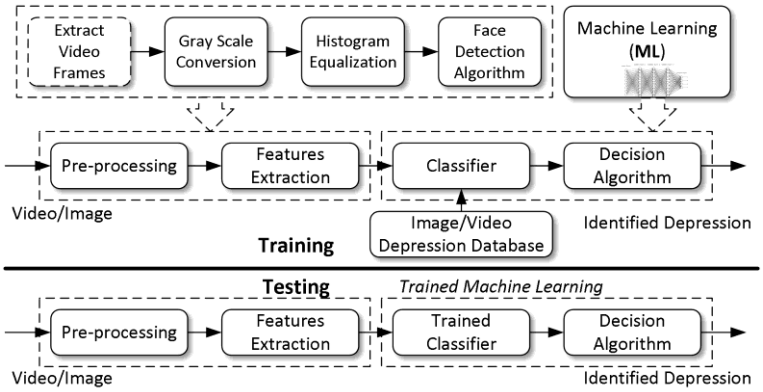

Figure 1. General machine learning architecture for depression detection

3.1. Preprocessing

Preprocessing is the first operation done on datasets to obtain the high level and useful data for the training and the testing phase. If the image/video is given input to our model, then all the normalization, illumination, alignment, segmentation, face detection, will be done in the preprocessing phase. There are many preprocessing tools enabling researchers to acquire intelligent information. Some of the best tools are the Computer Expression Recognition Toolbox, SEMAINE API, OpenFace freeware application [15-20]. The datasets are then categorized into training sets, which are used to train a machine learning model, and another set is called the testing set, which is used in the testing phase to check the performance of the model.

3.2. Feature extraction

It is an essential step in the process of training as well as testing because the rest of the steps further depend upon it. In the case of depression detection, features can be extracted from the face, head position, eyes, mouth [21]. PCA is considered most famous in this category [22-30]. However, to make a well-fitted system, these features have to be chosen very wisely. If the features selected will be less, then our model will face the problem of underfitting, and if too many features are selected in the training phase, the problem of overfitting may arise. Depending on different researchers, the numbers of features selected are different.

4. DEPRESSION INDICATORS

For a machine learning model to perform up to the merit, the model should have access to the comprehensive data from images, videos, voice, and linguistics. This section provides a review of the different depression indicators like visual indicators, voice indicators, and social indicators.

4.1. Visual indicators

Visual Indicators have broadly been explored for the detection of depression. Many researchers have conducted their work on finding a relation between non-verbal behavior and depression. Girard et al. [31] explored whether a relationship existed between nonverbal conduct and depression. To quantify nonverbal conduct, they utilized the Facial Action Coding Framework (FACS). FACS is a framework used to taxonomize human facial developments by their appearance on the face. It is a generally utilized tool and has become standard to methodically classify physical articulations, which has demonstrated exceptionally helpful for psychologists. FACS consists of facial action units that represent the basic actions of an individual's muscles [32]. In addition to FACS for video investigation, others have considered Space-Time Interest Points (STIP) highlights [33], which catch spatio-fleeting changes, including movements of the face, hands, shoulder, and head. Utilizing STIP highlights, they found that they could recognize depression with 76.7% precision.

4.2. Speech indicators

Recent researches have shown great confidence in using speech as the diagnostic tool for detecting depression. The speech mechanism of an individual is very complex and can be affected by the psychological state of mind. Speech can provide a tremendous amount of help in detecting the mood of any individual [34]. Many experts have done their work in determining the relation between the speech and the depression. Experts analyzed the relationship between depression and speech by performing factual investigations of various acoustic measures, including talking rate, percentage of delay time, and pitch variety. Their outcomes illustrated that talking rate and pitch variety had a solid relationship with the depression rating scale.

Scherer et al. [35] found glottal features contrasted fundamentally between depressed and normal groups. At the point when used to identify depression, they found glottal highlights to separate between the two groups with 75% precision. They examined various feature sets for distinguishing depression from unconstrained speech and discovered loudness and intensity features to be the most discriminative. According to the technique used by [36], the Speech Emotion Recognition (SER) rate over five emotions, namely happy, angry, sad, fear, and neutral, achieved was with 91.7% accuracy by introducing Voice Activity Detection (VAD) in preprocessing. For creating a model based on speech indicators, it is vital to use a formalized dataset. According to [37], while assessing an emotion recognition system, one of the significant factors to be thought of is the degree of instinctive nature of the database used to check the recognition execution.

4.3. Social indicators

Apart from visual and speech analysis, some experts carried out their work in detecting depression on social indicators. It is observed that a depressed person's body language is weak and has negative social behavior. Many such features can be used to diagnose depression. The social impacts of depression change how individual capacities on the planet and their association with others. Social impacts of sorrow include substance use and misuse, social and family withdrawal, and diminished execution.

Resnik et al. [38] investigated the utilization of administered point models in the investigation of depression from Twitter. Althoff et al. [39] displayed an enormous scale quantitative examination on the talk of directing discussions. They created many features to quantify how associated semantic parts of discussions were with results. Highlights in their examination included: arrangement based discussion models, language model examinations, message bunching, and psycholinguistics-enlivened word recurrence investigations. Xinyu Wang et al. developed a model to identify depression from online blog entries with a precision of 80% in terms of classification between depressed and non-depressed [40].

5. COMPARISON OF VARIOUS DEPRESSION DETECTION MODELS

The principal evaluation criteria for the detection of depression according to many kinds of research have been: presence (detected or not detected) and level prediction (low, mild, severe). These tasks involve evaluation metrics to report classification/prediction in terms of accuracy, precision, sensitivity, or harmonic mean of precision and recall (F1-score). These evaluation criteria are severely affected by the skewness in datasets used in the development of the model. While building a detection model, many factors have to be kept in mind like age, gender, and emotional behavior. Based on evaluation metrics given by Morales et al. [32], the best depression detection systems based on performance parameters (Accuracy, Mean Absolute Error (MAE) and Precision) are shown in Table 2, Table 3 and Table 4.

Table 2. Best performing depression detection models based on the accuracy

Reference	Features	Accuracy
Fraser et al. [41]	Mel-frequency cepstral coefficients	66%
Milne et al. [42]	N-grams	78%
Kim et al. [43]	TF-IDF n-gram/post embedding	85%
Malmasi et al. [44]	Lexical/syntax/metadata	83%
Brew [45]	TF-IDF unigrams/metadata	79%

Table 3. Best performing depression detection models based on MAE

Reference	Features	MAE
Valstar et al. [46]	[Visual, Acoustic, All]	[6.12, 5.72, 5.66]
Yang et al. [47]	Visual/acoustic	6.70
Williamson et al. [48]	[Visual, Acoustic, Semantic, All]	[5.33, 5.32, 3.34, 4.18]

Table 4. Best performing depression detection models based on precision

Reference	Features	Precision
Valstar et al. [46]	[Visual, Acoustic, All]	[0.47, 0.27, 0.47]
Yang et al. [47]	Visual/acoustic	0.50

Studies have revealed that to build an optimized model for the detection of depression, the data should be collected from a diverse population. More emphasis and more priority are given to the data collected from young participants. The data from combined sources proved to be less accurate, involving participants from older age groups [41]. Data from the young population needs to be considered to overcome this problem. The accuracy and precision evaluation parameters can also be computed using the confusion matrix. The "True Positive, True Negative, False Positive, and False Negative" can be computed [49], as shown in Figure 2.

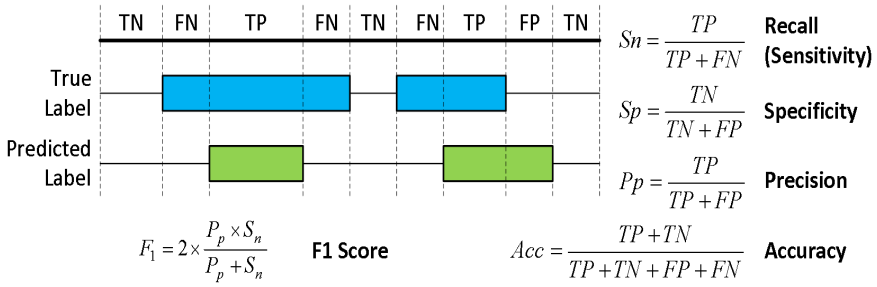

Figure 2. Performance metrics in machine learning

Multimodal machine learning techniques can be used to increase the performance parameters. In video-based detection systems, more video samples of an individual can be taken at various instances to analyze past and present mental state. Testing videos with more frames can be used for improvement. Periodic time evaluation can further be added to the approach to analyze the long term state of an individual. Apart from taking ready to use data for building the model more labeled data from the psychiatrists and mental health specialists can be taken to obtain more accurate results. Deep learning algorithms with deep architecture can be used for model development using image/video features.

6. EVALUATION OF MACHINE LEARNING ALGORITHMS

Machine learning techniques have entirely revolutionized the prediction and classification domain in terms of emotion recognition and depression detection. There are different machine learning techniques that have been incorporated to obtain the desired model to predict depression with more accurate performance. Different techniques of machine learning are used by different researchers like SVM, Neural Networks, Deep Learning. A support vector machine is a technique of machine learning which can be used for regression and classification tasks. However, most widely, it is used for classification tasks. The main objective of the support vector machine is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. In terms of emotion recognition, the SVM classifier is trained with different emotional features extracted in the training phase. The features extracted from each frame can be tested with the SVM classifier to check if these features of emotion are present in the frame. It only classifies based on features present in the frame, which means if there are fewer features of any emotion in the particular frame, then the intensity of that emotion is also less. Likewise, if the intensity of features in a particular frame is high, then the emotion is considered as high based on the hyperplane.

According to [4], videos from 5 different students were taken for experimental analysis. Each frame of every video was examined based on depression determined as positive or negative, i.e., +1 or -1. Overall, 160 images of the test video were considered. Around 65 video frames were correctly classified as having negative emotion, and the remaining 16 frames incorrectly classified as the positive class. For the positive emotion frames, out of the 79 frames, 38 were correctly classified as positive, and the remaining 41 were wrongly classified [4]. The performance metrics of the system for a video sample is shown in Table 5 and 6.

Table 5. Confusion matrix for a video sample [4]

Emotion	Negative (Actual)	Positive (Actual)	Total
Negative (Predicted)	65	16	81
Positive (Predicted)	41	38	79
Total	106	54	160

Table 6. Performance metrics of the system for a sample video [4]

Performance Metric	Value (%)
Accuracy	64.38
Error	35.62
Sensitivity	61.32
Specificity	70.37
Precision	80.25
False Positive Rate	29.63
F1_score	69.52

Other machine learning techniques, like neural networks or deep learning algorithms, are also passionately explored in the field of depression prediction from images/videos. The types of neural networks include Feed Forward Neural Network – Artificial Neuron, Radial Basis Function Neural Network, Kohonen Self Organizing Neural Network, Recurrent Neural Network (RNN) – Long Short Term Memory, Convolutional Neural Network (CNN), Modular Neural Network [50], Residual Depthwise Separable CNN [51-52]. Some unsupervised ML techniques like K-means can be used to determine the depression from the facial images. According to [53], the chance of utilizing facial pictures to decide whether a non-depressed individual is probably going to develop depression within 1-2 years. The prediction technique utilizes a run of the mill arrangement approach of preparing and testing. Class models were resolved utilizing picture information from teenagers who were either "in danger" or "not in danger" of depression. Two feature extraction approaches were looked at; the eigenface (PCA) features and the Fisherface (PCA+LDA) features. The nearest neighbor (NN) arrangement was actualized utilizing individual dependent and individual independent approaches. Best outcomes were with the Fisherface (PCA+LDA) strategy, giving expectation precision of 51% with the unique independent methodology and 61% when utilizing an individual dependent approach.

Contingent upon the specific research objectives, various sorts of machine learning decision techniques might be applied. Classification techniques are fitting to address clear cut inquiries (e.g., "depressed" versus "non-depressed" and low versus high depression severity). At the point when the exploration question concerns the simultaneous expectation of depression seriousness through video-determined files in a nonstop way, regression approaches are transcendentally utilized. Cross-validation techniques are typically applied before the classification/regression step.

7. CONCLUSIONS

This paper reviews several machine learning techniques for depression detection. This paper takes into consideration the fact that there are significant differences in the non-verbal and verbal actions between a depressed and non-depressed person. The data collection and data manipulation for building a detection model have been reviewed. Different machine learning techniques and tools used by different experts have been reviewed, along with the database used. The best performance parameters have been revived along with the techniques to determine the benchmarks. Depression is the most prevalent mental disorder present in our society at present, and there is a need to detect that for proper assessment. Many models have been developed for the cause, each having its advantages over the other. However, there is a need for an optimized and more validated detection model. These systems are not implemented as standalone diagnostics, but they can be beneficial for remote assessment, awareness, and support systems. This review provides several insights and also identifies many questions open to further investigation. To summit up all, it is essential to state that the image/video-based depression detection systems using machine learning techniques are valuable for both research and clinical practice to achieve better results towards raising the standard of living.

ACKNOWLEDGMENTS

The author would like to express their gratitude to the Malaysian Ministry of Education (MOE), which has provided research funding through the Fundamental Research Grant, FRGS19-076-0684. The authors would also like to thank International Islamic University Malaysia (IIUM), Universiti Teknologi MARA (UiTM) and Universitas Potensi Utama for providing facilities to support the research work.

REFERENCES

[1] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M.C. Gilardi, and A. Quattrone, "Machine Learning on Brain MRI Data for Differential Diagnosis of Parkinson's Disease and Progressive Supranuclear Palsy," *Journal of Neuroscience Methods*, 222, pp. 230-37, 2014.

[2] W. Katon and M.D. Sullivan, "Depression and chronic medical illness," *Journal of Clinical Psychiatry*, 51 (Suppl6), pp. 3-11, 1990.

[3] A.H. Orabi, P. Buddhitha, M.H. Orabi, and D. Inkpen, "Deep Learning for Depression Detection of Twitter Users," Proc. 5th Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pp. 88-97, 2018.

[4] N.S. Parameswaran and D. Venkataraman, "A computer vision based image processing system for depression detection among students for counseling," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 14, no. 1, pp. 503-512, 2019.

[5] A. Pampouchidou, P. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis, and M.Tsiknakis, "Automatic Assessment of Depression Based on Visual Cues: A Systematic Review," *IEEE Transactions on Affective Computing*, vol. 10, no. 4, pp. 445-470, 2019.

[6] Media.news.health.ufl.edu. [online] Available at: http://media.news.health.ufl.edu/misc/bolt/Intro/files/unit1/depression.xls [Accessed 3 Feb. 2020].

[7] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "AVEC 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM International Workshop on Audio/visual Emotion Challenge*, pp. 3-10, 2013.

[8] 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*, pp. 3-10, 2014.

[9] G.G.J. Chen, *Visualizations for mental health topic models*, Doctoral Dissertation, Massachusetts Institute of Technology, 2014.

[10] J. Gratch, R. Artstein, G.M. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, and D.R. Traum, "The distress analysis interview corpus of human and computer interviews," in *LREC*, pp. 3123-3128, 2014.

[11] J.T. Becker, F. Boiler, O.L. Lopez, J. Saxton, and K.L. McGonigle, "The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis," *Archives of Neurology*, vol. 51, no. 6, pp. 585-594, 1994.

[12] D.N. Milne, G. Pink, B. Hachey, and R.A. Calvo, "Clpsych 2016 shared task: Triaging content in online peer-support forums," In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pp. 118-127, 2016.

[13] S. Pradhan, W. Chapman, S. Man, and G. Savova, "Semeval-2014 task 7: Analysis of clinical text," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, vol. 199, pp. 54-62, 2014.

[14] H. Keshavarz, D. Fitzpatrick-Lewis, D.L. Streiner, R. Maureen, U. Ali, H.S. Shannon, and P. Raina, "Screening for depression: a systematic review and meta-analysis," *CMAJ open*, vol. 1, no.4, p. E159, 2013.

[15] T.Baltrušaitis, P. Robinson, and L.P. Morency, "Openface: an open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision*, pp. 1-10, 2016.

[16] M. Schröder, "The SEMAINE API: towards a standards-based framework for building emotion-oriented systems," *Advances in Human-Computer Interaction*, 2010.

[17] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (CERT)," in *Face and Gesture 2011*, pp. 298-305, 2011.

[18] K. A. Funes Mora, L. Nguyen, D. Gatica-Perez, and J.-M. Odobez, "A Semi-automated System for Accurate Gaze Coding in Natural Dyadic Interactions," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ser. ICMI '13. New York, NY, USA: ACM, 2013, pp. 87-90.

[19] K. A. Funes-Mora and J.-M. Odobez, "Gaze Estimation in the 3D Space Using RGB-D Sensors," *International Journal of Computer Vision*, vol. 118, no. 2, pp. 194-216, 2016.

[20] L.A. Jeni, J.F. Cohn, and T. Kanade, "Dense 3D Face Alignment From 2D Videos in Real-time," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, May 2015, pp. 1-8.

[21] C.A. Corneanu, M.O. Simón, J.F. Cohn and S.E. Guerrero, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548-1568, 2016.

[22] J.F. Cohn, T.S. Kruez, I. Matthews, Y. Yang, M.H. Nguyen, M.T. Padilla, F. Zhou, and F. De La Torre, "Detecting Depression from Facial Actions and Vocal Prosody," in *IEEE 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1-7, 2009.

[23] J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, and D.D. Mehta, "Vocal and Facial Biomarkers of Depression Based on Motor Incoordination and Timing," in *4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14)*, pp. 65-72, 2014.

[24] A. Jan, H. Meng, Y. F. A. Gaus, F. Zhang, and S. Turabzadeh, "Automatic Depression Scale Prediction using Facial Expression Dynamics and Regression," in *4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14)*, pp. 73-80, 2014.

[25] M. Senoussaoui, M. Sarria-Paja, J.F. Santos, and T.H. Falk, "Model Fusion for Multimodal Depression Classification and Level Detection," in *4th ACM International Workshop on Audio/ Visual Emotion Challenge (AVEC '14)*, pp. 57-63, 2014.

[26] C. Smailis, N. Sarafianos, T. Giannakopoulos, and S. Perantonis, "Fusing Active Orientation Models and Mid-term Audio Features for Automatic Depression Estimation," in *9th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2016.
[27] V. Jain, J.L. Crowley, A.K. Dey, and A. Lux, "Depression Estimation Using Audiovisual Features and Fisher Vector Encoding," in *4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14)*, pp. 87-91, 2014.
[28] H. Kaya, F. Çilli, and A.A. Salah, "Ensemble CCA for continuous emotion prediction," In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC'14)*, pp. 19-26, 2014.
[29] L. He, D. Jiang, and H. Sahli, "Multimodal Depression Recognition with Dynamic Visual and Audio Cues," in *IEEE International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 260-266, 2015.
[30] P. Wang, F. Barrett, E. Martin, M. Milonova, R. E. Gur, R. C. Gur, C. Kohler, and R. Verma, "Automated Video-based Facial Expression Analysis of Neuropsychiatric Disorders," *Journal of Neuroscience Methods*, 168(1), pp. 224-238, 2008.
[31] J.M. Girard, J.F. Cohn, M.H. Mahoor, S.M. Mavadati, Z. Hammal, and D.P. Rosenwald, "Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses," *Image and vision computing*, vol. 32, no. 10, pp. 641-647, 2014.
[32] M. Morales, S. Scherer, and R. Levitan, "A Cross-modal Review of Indicators for Depression Detection Systems," *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology-From Linguistic Signal to Clinical Reality*, pp. 1-12, 2017.
[33] N. Cummins, J. Joshi, A. Dhall, V. Sethu, R. Goecke, and J. Epps, "Diagnosis of depression by behavioural signals: a multimodal approach," In *Proc. of the 3rd ACM international workshop on Audio/visual emotion challenge*, pp. 11-20, 2013.
[34] M. Tasnim and E. Stroulia, "Detecting Depression from Voice," *Canadian Conference on Artificial Intelligence*, 2019.
[35] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morenc, "Investigating voice quality as a speaker-independent indicator of depression and PTSD," in *Interspeech*, pp. 847-851, 2013.
[36] M.F. Alghifari, T.S. Gunawan, M. Kartiwi, "Speech emotion recognition using deep feedforward neural network," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 10, no. 2, pp. 554-561, 2018.
[37] S.A.A. Qadri, T.S. Gunawan, M.F. Alghifari, H. Mansor, M. Kartiwi, and Z. Janin, "A critical insight into multi-languages speech emotion databases," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 8, no. 4, pp. 1312-1323, 2019.
[38] P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.A. Nguyen, and J.B.-Graber, "Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter," in *Proc. of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Singal to Clinical Reality*, pp. 99-107, 2015.
[39] T. Althoff, K. Clark, and J. Leskovec, "Large-scale analysis of counseling conversations: An application of natural language processing to mental health," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 463-476, 2016.
[40] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," In *Pacific-Asia Conference on Knowledge Discrovery and Data Mining*, pp. 201-213, 2013.
[41] KC. Fraser, F. Rudzicz, and G. Hirst, "Detecting late-life depression in Alzheimers disease through analysis of speech and language," *Proc. of the Third Workshop on Computational Lingusitics and Clinical Psychology*, 2016.
[42] D.N. Milne, G. Pink, B. Hachey, and R.A. Calvo, "Clpsych 2016 shared task: Triaging content in online peer-support forums," In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pp. 118-127, 2016.
[43] S. Mac Kim, Y. Wang, S. Wan, and C. Paris, "Data61-csiro systems at the clpsych 2016 shared task," In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pp. 128-132, 2016.
[44] S. Malmasi, M. Zampieri, and M. Dras, "Predicting post severity in mental health forums," In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pp. 133-137, 2016.
[45] C. Brew, "Classifying ReachOut posts with a radial basis function SVM," In *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology*, pp. 138-142, 2016.
[46] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M.T. Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, "AVEC 2016: Depression, mood, and emotion recognition workshop and challenge," In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, pp. 3-10, 2016.
[47] L. Yang, D. Jiang, L. He, E. Pei, M.C. Oveneke, and H. Sahli, "Decision tree based depression classification from audio video and language information," In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, pp. 89-96, 2016.
[48] J.R Williamson, E. Godoy, M. Cha, A. Schwarzentruber, P. Khorrami, Y. Gwon, H.-T. Kung, C. Dagli, and T.F. Quatieri, "Detecting depression using vocal, facial and semantic communication cues," In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, pp. 11-18, 2016.
[49] E. Victor, Z.M. Aghajan, A.R. Sewart, and R. Christian, "Detecting depression using framework combining deep multimodal neural networks with a purpose-built automated evaluation," *Psychological Assessment*, 2019.
[50] T.S. Gunawan, M.F. Alghifari, M.A. Morshidi, and M. Kartiwi, "A review on emotion recognition algorithms using speech analysis," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 6, no. 1, pp. 12-20, 2018.
[51] E. Ihsanto, K. Ramli, D. Sudiana, and T.S. Gunawan, "An Efficient Algorithm for Cardiac Arrhythmia Classification Using Ensemble of Depthwise Separable Convolutional Neural Networks," *MDPI Applied Sciences*, vol. 10, no. 2, 2020.
[52] E. Ihsanto, K. Ramli, D. Sudiana, and T.S. Gunawan, "Fast and Accurate Algorithm for ECG Authentication using Residual Depthwise Separable Convolutional Neural Networks," *MDPI Applied Sciences*, vol. 10, no. 9, 2020.
[53] K.E.B. Ooi, L.-S.A. Low, M. Lech, N. Allen, "Prediction of clinical depression in adolescents using facial image analysis," in *12th International Workshop on Image Analysis for Multimedia Interactive Services*, 2011.