IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 3, June 2025, pp. 2358∼2366
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i3.pp2358-2366

Depression and post traumatic stress disorder analysis with multi-modal data

Rajalakshmi Sivanaiah, Angel Deborah Suseelan, Krupa Elizabeth Thannickal, Sanmati Pirabahar
Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, India

Article Info

Article history:

Received Sep 3, 2024
Revised Feb 14, 2025
Accepted Mar 15, 2025

Keywords:

Artificial intelligence
Deep learning
Ensemble learning
Mental health
Text, video and audio data
Transformer models

ABSTRACT

With an increasing global population and more people living to the age when major depressive disorder (MDD) or post traumatic stress disorder (PTSD) commonly occurs, the number of those who suffer from such disorders is rising. Studies have also shown a high likelihood of comorbidity between these 2 disorders. This comorbidity can worsen symptoms, increase the risk of chronicity, and complicate treatment, significantly impacting patients' emotional well-being and social and occupational functioning. There is a need to enable faster and reliable diagnosis methods, while taking into account the subjectivity of individuals and the role of behavioural cues. The proposed approach analyses the combination of audio, video and text input features (multi-modal data) of the subject to determine the severity class of MDD and PTSD. The DistilBERT transformer is used for learning and building a model with the textual modality and random forest classifiers for the audio and video modalities. An ensemble of these 3 models from 3 modalities performs better in the final classification of MDD and PTSD when compared to individual models. This work also covers a comparison of the models with different splits on the dataset. This ensembled system shows an improved accuracy of 2% to 7% for the MDD and PTSD multi class classification over the models tested on individual modalities.

Corresponding Author:

Rajalakshmi Sivanaiah
Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering
Chennai - 603110, Tamil Nadu, India
Email: rajalakshmis@ssn.edu.in

1. INTRODUCTION

The role of machine learning and deep learning to automate disease diagnosis and treatment is ever-increasing. One such area where automation could be of immense help is in mental health diagnosis. According to the World Health Organization (WHO), mental health disorders are the major contributors to years lived with disability globally. The first step towards helping such patients is obviously the proper and correct identification of the disorder they might be suffering from in initial stage itself. But as per a WHO 2022 report [1], nearly 50% of the global population resides in countries where the ratio of psychiatrists to people is approximately one psychiatrist for every 200,000 individuals or more. The two disorders this paper focuses on are depression, also known as major depressive disorder (MDD) and post traumatic stress disorder (PTSD). Brewin *et al.* [2] have shown high correlation between MDD and PTSD, with approximately 50% of people with PTSD also suffering from MDD. Current methodology to assess depression and PTSD is using self-report questionnaires, like patient health questionnaire (PHQ) PHQ-8 and PTSD Check List - Civilian version (PCL-C) respectively, followed by clinical interviews and a review of medical history. This system relies on several variable factors and is not very reliable due to multiple subjective issues and behavioural cues of the individual. Also, present psychosocial interviews take longer time, while an automated system like Ellie, that was used to retrieve the Extended Distress Analysis Interview Corpus (E-DAIC) dataset, takes lesser time to engage the patient. As in the case of most medical treatment, an early diagnosis increases chances of a faster recovery and can prevent the onset of worse conditions. This is the most pressing reason to identify the co-existence of depression and PTSD as ongoing research suggests [2] a link between the two conditions could influence treatment approaches. One of the areas where the International Audio/Visual Emotion Challenge and Workshop (AVEC) has attempted to spur interest is in depression detection. In 2019, the E-DAIC dataset described in section 2.1. was provided along with the results of a baseline model to participants [3].

The baseline model in this challenge considered a binary classification for depression detection using a gated recurrent unit (GRU) - recurrent neural network (RNN) model. One of the more successful attempts from AVEC 2019 was of a multi-level attention network that used a fusion approach to select the most important features of each input modality [4]. These are given to a series of feed-forward networks whose final outputs are again fused in a stacked bidirectional long short-term memory (BLSTM) model. The resulting system succeeded in outperforming the challenge baseline root mean square error (RMSE) by 17.52%, while also identifying the relative importance of the different input data types. The study utilizes databases from multiple languages for acoustic feature selection and then extracts text-based features for depression assessment [5]. Given the small amount of training data available, effective data selection is crucial, and the proposed multi-lingual method outperformed baseline algorithms in feature selection, improving depression assessment accuracy. Research has also been done for a predictive model from text using a combination of a long-short term memory (LSTM) model and a RNN [6]. The RNN is trained on textual data to recognize depression based on semantics and written content. The proposed framework has a 99.0% accuracy rate, which is higher than frequency-based deep learning models for text-based detection, and also has a lower false positive rate. Zavorina and Makarov describe a transformer encoder network in their study on voice-base depression detection [7], compared to top baseline approaches using the E-DAIC dataset. To handle the small size of the available dataset, low-level features are extracted from audio recordings and then augmented. Their network attains a recognition accuracy of 73.51% on the E-DAIC database, which is better than the accuracy of 65.85% and 66.35% obtained by traditional approaches. Research on visual feature selection [8] shows that the high dimensionality of these features poses a challenge in identifying informative and discriminative ones. To address this, a hybrid dimensionality reduction approach is introduced that combines filter and wrapper methods.

Using the technique of mobile crowd sensing combined with a task-based approach, a design for an end-to-end ML pipeline from data collection to binary classification has also been proposed [9]. This study experimented with various features from multi-modalities, feature selection techniques, fused features, and machine learning classifiers such as logistic regression and support vector machines (SVM). The SVM algorithm produced the best accuracy, reaching 86%, when features from all three modalities were fused. The review on the need and the methods used for PTSD are discussed in [10], [11]. Apart from the PHQ-8 scale for depression, the Hamilton Depression Rating Scale (HDRS) is also used in research to train models for detecting depression severity [12]. This technique employs deep neural networks to detect video, speech, and text modalities, and combines the modalities using weighted fusion to calculate the total points in the HDRS. Results showed that the video modality achieved an accuracy of 66%, the speech modality achieved 81%, and 82% for the text modality. Apart from the E-DAIC and DAIC-WoZ dataset, studies have been carried out on the D-Vlog dataset [13], comprising 961 YouTube vlogs. This research developed models based on the non-verbal behavior of individuals in real-world scenarios, using only the acoustic and visual features. A cross-attention mechanism was used to discover the connection between features, followed by comparisons with baseline models.

There is a scanty of existing work available for depression detection system with multi-modal data. Either they use a unimodal [14] or bi-modal [15] data whereas we concentrate on extracting features from multi-modal data of a particular user for better understanding of the user's behaviour. Mostly all the works are done for binary classification (depression or not) and we have considered a multi-class scenario (5 classes to show the level of depression). The co-occurrence of both MDD and PTSD in an individual is also not analysed. We have aimed to concentrate on these issues and develop a multi-modal system to detect the depression and PTSD. This section explains the need and current research in depression and PTSD detection. Section 2 describes the dataset and methodologies used to solve the problem. Section 3 explains the experimental results and the discussions. Section 4 concludes the findings and future work.

2. METHOD

2.1. Dataset description

The E-DAIC dataset consists of 275 interviews of participants collected both in a Wizard of Oz mode (human-controlled agent) and with a purely AI agent named Ellie [16]. Each of the participant data labelled as XXX_P has been organized into a directory with raw audio files for each interview (XXX_Audio.wav). Tanscripts of each audio file (XXX_Transcript.csv), pose, gaze and action units from video collected with a time step of 0.033 seconds and PHQ-8 and PCL-C scores of participants. These scores are used to determine the severity class labels of the samples as described in subsection 2.2.

2.2. Modules

2.2.1. Cleaning, feature extraction and labelling

Previous studies have pointed out some known issues with the raw audio files in the E-DAIC dataset [17]. For example, there are long intermittent pauses in some interviews which should be removed. To handle these the raw audio is trimmed by removing regions having volumes less than 33dB. The OpenSMILE toolkit is used to extract the Compare2016 set of 65 features from the trimmed raw audio. A frame size of 0.1 second was used for this. Given the varying length of individual interviews, Compare2016 features and pose, gaze and action units for video are also variable in length for each participant. To deal with this, a simple averaging of the extracted features is done, so that this is reduced to a single feature frame each for audio and video. For each interview, the text transcript is split into timestamps as separate lines or phrases of conversation. All texts for a single participant are concatenated into one single row of text. The labelling was done for the cleaned sample based on the range in which the corresponding PHQ-8 and PCL-C score falls under. The ranges used for labelling are mentioned in Tables 1 and 2, adopted from the existing standards for MDD and PTSD classification [18], [19]. The dataset comprises of the PHQ-8 score and PCL-C score only and we have converted the score to multiple class labels based on the segregation values mentioned in Tables 1 and 2 using [20], [21]. Figure 1 shows the data distribution for the MDD dataset for training, development and test splits. The outer circle denotes the training dataset, middle circle denotes the development dataset and the inner circle denotes the test dataset. Figure 2 shows the data distribution for PTSD dataset for training, development and test datasets.

Table 1. MDD severity labels from PHQ-8 scores

PHQ-8 score	Severity class
0-4	Absence
5-9	Mild
10-14	Moderate
15-19	Moderate to severe
>=20	Severe

Table 2. PTSD severity labels from PCL-C scores

PCL-C score	Severity class
17-27	Absence
28-29	Maybe mild
30-44	Moderate to high
45-85	Severe

2.2.2. Multimodal data - individual models

After conducting several experiments with various machine learning (logistic regression, decision trees, random forest, SVM, k nearest neighbours, gaussian naive Bayes) [22] and deep learning models [23] (BERT [24], RNN, CNN, LSTM, Roberta, DistilBERT [25]) the final decision was made to choose the DistilBERT model for text and the support vector and random forest classifiers for audio and video data, as these models exhibited a higher accuracy rate compared to other models. The DistilBERT model used is a transformer model under Python's torch library, where the encoded input text is passed through a pre-trained "distilbert-base-uncased" model that generates a sequence of contextualized embeddings for each token in it. This is followed by a dropout layer having a dropout rate of 0.1 to reduce overfitting, and a linear classification layer that maps the hidden state size of the DistilBERT model [26] to the number of output classes. This produces the logits, which are raw scores indicating the likelihood of each input belonging to each output class.

The random forest classifier is used for MDD audio and video training and for PTSD video modelling [27]. The support vector classifier was chosen for PTSD audio training as it displayed a higher accuracy than the corresponding random forest model. The training was carried out for individual models under text, audio, and video data using the cleaned and labelled samples and saved to be used later for the ensemble. These saved models were loaded for ensemble learning, where the outputs from multiple models are combined to improve the overall prediction accuracy.

2.2.3. Multimodal data - ensemble models

For each participant of the testing samples, the final class is decided based on majority voting. The class that is given by 2 out of 3 modalities is set to be the label for that particular participant. If all three modalities resulted in different labels, the final label was set as an average of the numerical values corresponding to the labels. Figure 3 shows the overall architecture of the proposed system to detect MDD and PTSD using multimodal data in the form of audio, video and text input of the user. Machine learning models are built for the individual modalites of E-DAIC dataset and they are ensembled to produce the final output class label.

The first step of the system uses Python's librosa and soundfile libraries to trim parts of raw audio where decibel levels were less than 33 dB. From this trimmed audio, 65 features were extracted under the Compare2016 set using the OpenSMILE toolkit. These were derived using a frame size of 0.1 seconds. The set of video features used from the E-DAIC dataset were the 49 pose, gaze and action units. These had been dervied using the OpenFace toolkit with a frame size of 0.033 seconds. A simple averaging of the extracted features for both audio and video was then carried out for each of the participant sample files. The text transcript for each individual participant from their corresponding csv files was combined into one single row of text. It was observed that the training, testing and validation split provided in the E-DAIC dataset given in Figures 1 and 2 were not uniformly distributed across the new labels assigned for the multi-class approach. In order to compare performances, a new split up was created in the ratio of 70:15:15 for training, testing and validation. This new split is visualized in Figures 4 and 5 for MDD and PTSD respectively.

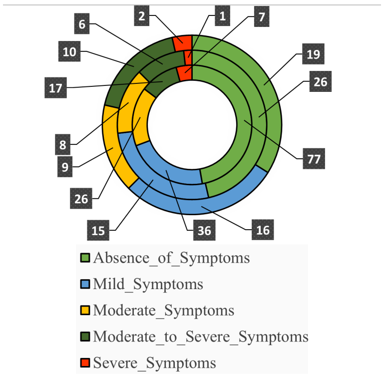

Figure 1. MDD multi-class distribution of E-DAIC dataset over train, dev and test splits

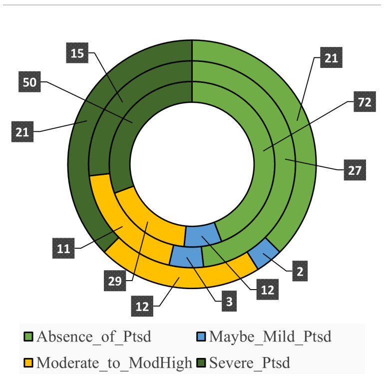

Figure 2. PTSD multi-class distribution of E-DAIC dataset over train, dev and test splits

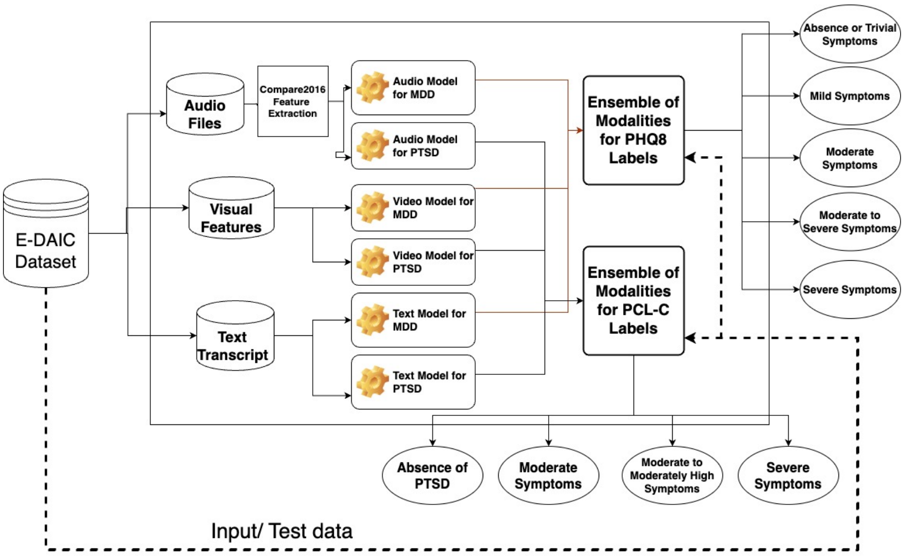

Figure 3. System architecture

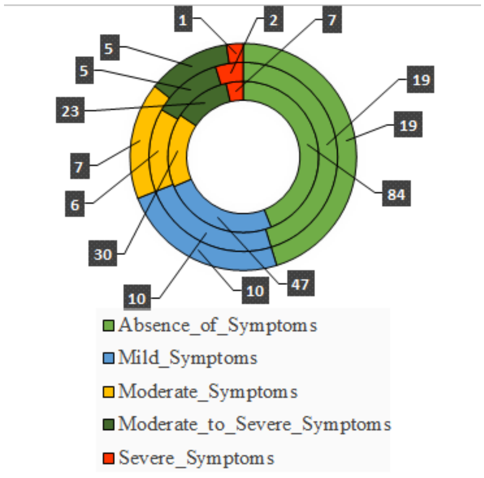

Figure 4. MDD multi-class distribution over new train, dev and test splits

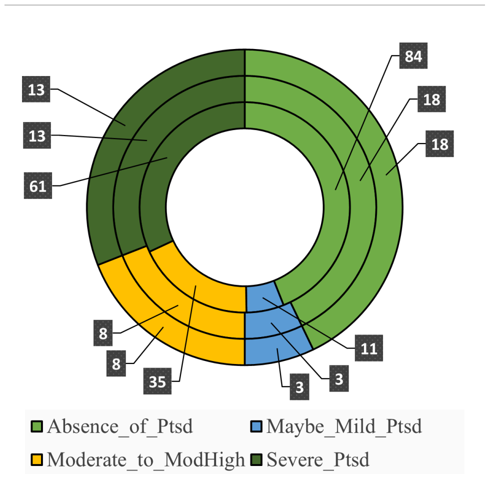

Figure 5. PTSD multi-class distribution over new train, dev and test splits

3. RESULTS AND DISCUSSION

This paper discusses the key findings on detecting the depression level and PTSD level as a multi-class problem, while the earlier studies deal this as a regression or binary-class problem. We have devised the rules for classifying the scores into multiple categories and experimented with the need for multiple modalaties in detecting the level of MDD and PTSD. The evaluation of the three trained models for individual modalities and the ensemble model for MDD and PTSD includes the calculation of accuracy and F1-score. The results for all four models and their performance measured using F1 score are presented in Tables 3 and 4 for MDD and PTSD respectively. Table 5 depicts the performance of the proposed ensemble approach compared to other algorithms. We have observed that the new split created shows better results than the given E-DAIC split in terms of performance. Our study demonstrates that the ensemble model with all three modalities (text, audio and video) gives better results than the individual models for both MDD and PTSD classification. Another performance metric that visualizes and summarizes the performance of the multi-class classification is a confusion matrix that shows the number of correctly classified and misclassified samples under each class. The matrices for MDD and PTSD classification are displayed in Tables 6 and 7 respectively.

Table 3. F1-Score and accuracy of MDD classification

Modality	F1 E-DAIC	F1- New	Acc E-DAIC	Acc New
Text	0.34	0.38	35.7	40.5
Audio	0.31	0.35	39.3	45.2
Video	0.25	0.29	35.7	45.8
Ensemble	0.38	0.42	42.9	47.8

Table 4. F1-Score and accuracy of PTSD classification

Modality	F1 E-DAIC	F1- New	Acc E-DAIC	Acc New
Text	0.42	0.35	42.9	40.5
Audio	0.21	0.43	37.5	47.6
Video	0.38	0.42	46.1	50.0
Ensemble	0.47	0.48	48.6	52.4

Table 5. Accuracy of different machine learning algorithms

Task	DT	RF	SVM	KNN	Ensemble-New
MDD	32	33	33	33	47.8
PTSD	32	32	37	33	52.4

Table 6. Confusion matrix for MDD ensemble

Labels	Abs	Mild	Mod	ModSev	Sev
Abs	17	2	0	0	0
Mild	9	7	0	0	0
Mod	6	3	0	0	0
ModSev	8	2	0	0	0
Sev	1	0	1	0	0

Table 7. Confusion matrix for PTSD ensemble

Labels	Abs	Mild	Mod	Sev
Abs	20	0	0	1
Mild	2	0	0	0
Mod	9	0	1	1
Sev	16	0	1	5

From the confusion matrix of MDD shown in Table 6, out of 19 "Absence" (Abs) labelled test data samples 17 are correctly classified and 2 are misclassified as "Mild" class. Similarly the classification is shown for other class labels. However, it is worth noting that neither of the two test data samples of "Severe" (Sev) class are correctly classified. One sample is misclassified as "Absence" and the other one as "Moderate" (Mod). Same scenario exists for "Moderate to severe" (ModSev) category. These false negative classification's need to be reduced in the case of depression detection. From the confusion matrix of PTSD shown in Table 7, one test data sample of "Absence" is misclassified as "Severe" while 16 test data samples of "Severe" are misclassified as "Absence". Both of these false negatives need attention. In our future work we are working on reducing these false negatives using boosting algorithms. Classwise accuracy from the confusion matrices are determined for the individual classes of both MDD and PTSD as shown in Tables 8 and 9 respectively.

Table 8. Accuracy for MDD multi-class labels

Class	Accuracy
Absence	89.47
Mild	43.75
Moderate	0.0
Moderate to Severe	0.0
Severe	0.0

Table 9. Accuracy for PTSD multi-class labels

Class	Accuracy
Absence	95.24
Mild	0.0
Moderate to high	9.09
Severe	22.73

4. CONCLUSION

From the training of these model, it was observed that initially all the samples are set to belong to the ‘Absence’ class in both MDD and PTSD. The reason for this might be that the model begins to distinguish between classes only in later epochs or stages of training. However, because of the dataset imbalance in a multi-class approach, the results fail to improve with training. This imbalance has resulted in few classes being identified while other classes are not recognized by the ensemble model. Another reason for this is derived from the comparison the performance of models on the E-DAIC split versus the new split up of data. Since the new split distributed more samples for training, the accuracy is seen to be a bit better for models trained on this division. Secondly, despite the extensive features extracted, the limited size of the dataset to 275 participants on the whole, hinders the accurate identification in a multi-class scenario. If the dataset could be augmented to provide more, and equal number of samples for all classes, it is possible that the results might improve. It is also possible that by identifying the most relevant features for a participant, models could be trained better, but this might require assistance from experts in the particular subject area. There might be scope for better results depending on the algorithms and approaches used. While audio features have been extracted using the Compare2016 set in the proposed system, feeding the raw audio directly into the model can allow the model to discover its own patterns and feature sets. Using other feature sets for audio training could also yield better results. On the textual front, the varying length of transcripts can influence the model and how much of the actual content it focuses on. Choosing appropriate weights, or cleaning the dataset of irrelevant text, identifying spelling errors, grammatical mistakes, and other inconsistencies can help to improve the performance. We hope that our system will thereby enable further research into the comorbidity spectrum of these two disorders as this can directly impact medical research and related treatments. Early detection of the stress using multi-modal data and prescribing for treatments can reduce the casuality of MDD and PTSD.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rajalakshmi Sivanaiah	✓	✓		✓	✓	✓	✓		✓	✓		✓	✓	
Angel Deborah Suseelan	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	
Krupa Elizabeth Thannickal		✓	✓	✓		✓	✓		✓		✓			
Sanmati Pirabahar		✓	✓	✓		✓	✓		✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request and signing an agreement form present in the website https://dcapswoz.ict.usc.edu/.

REFERENCES

[1] M. Freeman, "The world mental health report: transforming mental health for all," *World Psychiatry*, vol. 21, no. 3, pp. 391–392, Oct. 2022, doi: 10.1002/wps.21018.

[2] C. R. Brewin, L. Atwoli, J. I. Bisson, S. Galea, K. Koenen, and R. Lewis-Fernández, "Post-traumatic stress disorder: evolving conceptualization and evidence, and future research directions," *World Psychiatry*, vol. 24, no. 1, pp. 52–80, Feb. 2025, doi: 10.1002/wps.21269.

[3] F. Ringeval *et al.*, "AVEC 2019 workshop and challenge: State-of-mind, detecting depression with AI, and cross-cultural affect recognition," *AVEC 2019 - Proceedings of the 9th International Audio/Visual Emotion Challenge and Workshop, co-located with MM 2019*, pp. 3–12, Jul. 2019, doi: 10.1145/3347320.3357688.

[4] A. Ray, S. Kumar, R. Reddy, P. Mukherjee, and R. Garg, "Multi-level attention network using text, audio and video for depression prediction," *AVEC 2019 - Proceedings of the 9th International Audio/Visual Emotion Challenge and Workshop, co-located with MM 2019*, pp. 81–88, Sep. 2019, doi: 10.1145/3347320.3357697.

[5] C. Demiroglu, A. Beşirli, Y. Ozkanca, and S. Çelik, "Depression-level assessment from multi-lingual conversational speech data using acoustic and text features," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, Dec. 2020, doi: 10.1186/s13636-020-00182-4.

[6] A. Amanat *et al.*, "Deep learning for depression detection from textual data," *Electronics*, vol. 11, no. 5, Feb. 2022, doi: 10.3390/electronics11050676.

[7] E. Zavorina and I. Makarov, "Depression detection by person's voice," in *Analysis of Images, Social Networks and Texts*, 2022, pp. 250–262, doi: 10.1007/978-3-031-16500-9_21.

[8] S. Rathi, B. Kaur, and R. K. Agrawal, "Selection of relevant visual feature sets for enhanced depression detection using incremental linear discriminant analysis," *Multimed Tools Applications*, vol. 81, no. 13, 2022, doi: 10.1007/s11042-022-12420-2.

[9] R. P. Thati, A. S. Dhadwal, P. Kumar, and S. P, "A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms," *Multimed Tools Applications*, vol. 82, no. 4, pp. 4787–4820, 2023, doi: 10.1007/s11042-022-12315-2.

[10] D. Nilaweera *et al.*, "Lifetime posttraumatic stress disorder as a predictor of mortality: a systematic review and meta-analysis," *BMC Psychiatry*, vol. 23, no. 1, Apr. 2023, doi: 10.1186/s12888-023-04716-w.

[11] G. I. Al Jowf, Z. T. Ahmed, R. A. Reijnders, L. de Nijs, and L. M. T. Eijssen, "To predict, prevent, and manage post-traumatic stress disorder (PTSD): A review of pathophysiology, treatment, and biomarkers," *International Journal of Molecular Sciences*, vol. 24, no. 6, Mar. 2023, doi: 10.3390/ijms24065238.

[12] H. V. M. Sudhan and S. S. Kumar, "Multimodal depression severity detection using deep neural networks and depression assessment scale," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, vol. 99, 2022, pp. 361–375, doi: 10.1007/978-981-16-7182-1_29.

[13] J. Yoon, C. Kang, S. Kim, and J. Han, "D-vlog: Multimodal vlog dataset for depression detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 12226–12234, Jun. 2022, doi: 10.1609/aaai.v36i11.21483.

[14] S. Hemtanon, S. Aekwarangkoon, and N. Kittiphattanabawon, "Proactive depression detection from Facebook text and behavior data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 5027-5035, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5027-5035.

[15] A. Ashraf, T. S. Gunawan, B. S. Riza, E. V. Haryanto, and Z. Janin, "On the review of image and video-based depression detection using machine learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 3, pp. 1677–1684, 2020, doi: 10.11591/ijeecs.v19.i3.pp1677-1684.

[16] D. DeVault *et al.*, "SimSensei kiosk: A virtual human interviewer for healthcare decision support," in *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, 2014, pp. 1061–1068.

[17] A. Bailey and M. D. Plumbley, "Gender bias in depression detection using audio features," in *2021 29th European Signal Processing Conference (EUSIPCO)*, IEEE, Aug. 2021, pp. 596–600, doi: 10.23919/EUSIPCO54536.2021.9615933.

[18] I. Moshe *et al.*, "Digital interventions for the treatment of depression: A meta-analytic review," *Psychological Bulletin*, vol. 147, no. 8, pp. 749–786, 2021, doi: 10.1037/bul0000334.

[19] A. J. Angleman, V. B. Van Hasselt, and B. B. Schuhmann, "Relationship between posttraumatic stress symptoms and cardiovascular disease risk in firefighters," *Behavior Modification*, vol. 46, no. 2, pp. 321–351, 2022, doi: 10.1177/01454455211061320.

[20] J. Gratch *et al.*, "The distress analysis interview corpus of human and computer interviews," in *Proceedings of the 9th International Conference on Language Resources and Evaluation*, LREC 2014, 2014, pp. 3123–3128.

[21] I. A. Mureșanu *et al.*, "Evaluation of post-traumatic stress disorder (PTSD) and related comorbidities in clinical studies," *Journal of Medicine and Life*, vol. 15, no. 4, pp. 436–442, Apr. 2022, doi: 10.25122/jml-2022-0120.

[22] V. S. Sathvika, S. Vaishnavi, S. A. Deborah, S. Rajalakshmi, and T. T. Mirnalinee, "The Mavericks@LT-EDI-2023: Detection of signs of depression from social media texts using navie Bayse approach," in *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, 2023, pp. 244–249.

[23] K. Anantharaman, S. Rajalakshmi, S. A. Deborah, M. Saritha, and R. S. Milton, "SSN_MLRG1@LT-EDI-ACL2022: Multi-class classification using BERT models for detecting depression signs from social media Text," in *LTEDI 2022 - 2nd Workshop on Language Technology for Equality, Diversity and Inclusion, Proceedings of the Workshop*, 2022, pp. 296–300, doi: 10.18653/v1/2022.ltedi-1.44.

[24] V. O. Yenumulapalli, R. V. Aravindh, S. Rajalakshmi, and S. A. Deborah, "TechSSN1@LT-EDI: Depression detection and classification using BERT model for social media texts," in *LTEDI 2023 - 3rd Workshop on Language Technology for Equality, Diversity and Inclusion, associated with the 14th International Conference on Recent Advances in Natural Language Processing, RANLP 2023 - Proceedings*, 2023, pp. 149–154, doi: 10.26615/978-954-452-084-7_022.

[25] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” *arXiv-Computer Science*, pp. 1-5, 2019.
[26] K. E. Thannickal, P. Sanmati, S. Rajalakshmi, and S. A. Deborah, “TechSSN4@LT-EDI-RANLP2023: Depression sign detection in social media postings using DistilBERT model,” in *LTEDI 2023 - 3rd Workshop on Language Technology for Equality, Diversity and Inclusion, associated with the 14th International Conference on Recent Advances in Natural Language Processing, RANLP 2023*, 2023, pp. 239–243, doi: 10.26615/978-954-452-084-7_036.
[27] S. P. Pimpalkar, A. A. Chavan, S. V. Gawali, and S. S. Dalvi, “Advancements in depression severity prediction: A multi-modal machine learning approach integrating MCQs and audio data,” in *2nd IEEE International Conference on Advances in Information Technology, ICAIT 2024 - Proceedings*, 2024, pp. 1–8, doi: 10.1109/ICAIT61638.2024.10690826.

BIOGRAPHIES OF AUTHORS

Rajalakshmi Sivanaiah is working as Assistant Professor in Department of Computer Science and Engineering, SSN College of Engineering, Chennai, India. She has more than 19 years of experience in teaching. Her area of interest includes data mining, machine learning, deep learning, data science, natural language processing, and internet technologies. She completed Ph.D in recommendation systems under Anna University, Chennai, India. She is consistently taking part in the International Workshop on “Semantic evaluation (SemEval)” organized by Association of Computational Linguistics from 2018, MediaEval and LT-EDI tasks. She has published papers in various conferences and journals and granted with 1 patent. She has completed an internally funded project on fine grained emotion detection. She can be contacted at email: rajalakshmis@ssn.edu.in.

Angel Deborah Suseelan is Assistant Professor in the Department of Computer Science has 11 years and 6 months of teaching and research experience. She received her Ph.D. from Anna University. She is consistently taking part in the SemEval, MediaEval, and LT-EDI tasks. She has published papers in various conferences and journals and granted with 1 patent. Her research areas includes machine learning, natural language processing, data science, embedded systems and internet of things. She has completed an internally funded project on fine grained emotion detection. She can be contacted at email: angeldeborahs@ssn.edu.in.

Krupa Elizabeth Thannickal graduated with a Bachelor of Engineering (B.E) degree in Computer Science and Engineering. Currently employed in the capacity of Associate Software Engineer at Informatica, she applies her technical skills in developing software for process integration and data management. Her areas of interest include programming in Python, cryptography, and graphic design. She can be contacted at email: krupa19054@cse.ssn.edu.in.

Sanmati Pirabahar is a skilled professional with a Bachelor’s degree in Computer Science from SSN College of Engineering, passionate about operating systems, Java programming, microprocessors, and backend development. Currently employed as a Software Engineer at Informatica, she applies her expertise and enthusiasm to contribute effectively in various technological projects. With a keen interest in exploring advancements in these areas, she remains dedicated to enhancing her skills and driving innovation in the field of technology. She can be contacted at email: sanmati19098@cse.ssn.edu.in.