Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

DEPRESSION DETECTION USING DEEP LEARNING

Mr. K. Hari Veera Raju[1], Paleru Uma Naga Durga[2], Yadla Bhargavi[3], Ravada E L Priyanka[4], Shaik Rizwana[5]
#1Assistant Professor in Department of CSE, Raghu Engineering College, Dakamarri, Visakhapatnam.
#2#3#4#5 B.Tech with Specialization of Computer Science and Engineering in Raghu Engineering College, Dakamarri, Visakhapatnam.

ABSTRACT:
In the medical domain, explaining the reasoning behind model results is crucial for building trust and understanding predictions. This is concerning, especially in tasks like automatic depression prediction, where models often produce obscure predictions. Depression is a prevalent mental health condition that affects millions of people worldwide. Early detection and intervention are crucial for effective treatment. In this project, we propose a deep learning-based approach for automatic depression detection. We leverage the power of deep learning and machine learning algorithms to analyze user posts and identify patterns indicative of depression. This model incorporates multi-aspect features and employs hierarchical attention mechanisms to capture important information.This project contributes to the development of automated tools for mental health screening and has the potential to assist healthcare professionals in identifying individuals at risk of depression.
This paper presents an integrative deep learning approach for depression detection, leveraging the power of Support Vector Machine (SVM), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Random Forest (RF) algorithms. Depression remains a significant global health concern, and early detection is crucial for effective intervention. Traditional methods often lack the accuracy and efficiency required for timely diagnosis, hence the need for advanced computational techniques.
The proposed framework integrates multiple machine learning algorithms to analyze textual and visual data for depression symptoms. For textual analysis, a SVM model is employed to classify sentiment and extract relevant features from social media posts and other textual sources. Simultaneously, a CNN model is utilized for visual analysis, extracting features from images associated with depression, capturing subtle cues and patterns indicative of mental health status and LSTM networks are employed for processing sequential data such as textual and temporal signals. Additionally, a Random Forest ensemble model is employed to fuse the outputs of the SVM and CNN models, enhancing overall classification performance.To address the class imbalance and enhance model generalization and ensemble learning strategies are employed.
Experimental results on a large-scale dataset demonstrate the effectiveness of the integrative approach, showcasing superior performance compared to individual algorithms. The ensemble model achieves high accuracy, sensitivity, and specificity in depression detection, outperforming existing methods.
Overall, this research contributes to advancing depression detection methodologies by integrating diverse machine learning techniques, paving the way for more accurate and accessible mental health screening tools.

Index terms
Depression detection, Deep learning, Support Vector Machine (SVM), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Random Forest (RF), Textual data processing, Computer vision, Ensemble learning, Mental health screening, Real-world deployment

1.INTRODUCTION

In contemporary society, depression has emerged as a significant mental health concern, affecting millions worldwide. The complexity of diagnosing depression lies in its multifaceted nature,

characterized by subtle nuances and diverse manifestations across individuals. While traditional diagnostic methods rely heavily on subjective assessments and self-reporting, the integration of advanced computational techniques presents a promising avenue for enhancing detection accuracy and efficiency.

This paper proposes a comprehensive investigation into depression detection using a fusion of deep learning and classical machine learning algorithms, including Support Vector Machines (SVM), Random Forest (RF), and Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM). By leveraging the power of artificial intelligence, we aim to develop a robust and reliable framework capable of identifying depression symptoms with high precision and sensitivity.

Deep learning, characterized by its hierarchical architecture and ability to automatically extract intricate features from raw data, offers a compelling solution for analyzing complex mental health patterns. Convolutional Neural Networks, in particular, excel in image and signal processing tasks, making them well-suited for interpreting neuroimaging data and physiological signals associated with depression. Through the utilization of CNNs, we endeavor to discern subtle patterns and biomarkers indicative of depressive states, thereby enhancing diagnostic accuracy.

Complementing deep learning, classical machine learning algorithms such as SVM and RF provide a versatile toolkit for modeling and classification tasks. SVM, known for its effectiveness in handling high-dimensional data and nonlinear relationships, offers a principled approach to delineating depression-related patterns in heterogeneous datasets. Similarly, RF, with its ensemble of decision trees and robustness to overfitting, holds promise in capturing the complex interactions between various risk factors and depressive symptoms.

Central to our methodology is the integration of cross-validation techniques, ensuring the generalizability and reproducibility of our models across diverse populations and datasets. By systematically partitioning the data into training, validation, and test sets, we aim to mitigate the risk of overfitting and assess the robustness of our algorithms under different conditions.

The significance of our research extends beyond the realm of academic inquiry, with profound implications for clinical practice and public health policy. By providing clinicians with a reliable computational tool for early detection and intervention, we aspire to reduce the burden of undiagnosed and untreated depression, ultimately improving patient outcomes and quality of life.

In summary, this paper embarks on a journey to harness the collective power of deep learning and classical machine learning algorithms for the purpose of depression detection. Through a synergistic integration of advanced computational techniques and clinical insights, we strive to illuminate the path towards more effective, efficient, and equitable mental health care delivery.

2.LITERATURE SURVEY

1. Deep Learning Approaches for Depression Detection:

Deep learning has garnered significant attention in recent years for its potential in detecting depression from various modalities of data, including neuroimaging, speech, and text. Some studies have utilized Convolutional Neural Networks (CNNs) to analyze functional Magnetic Resonance Imaging (fMRI) data, identifying distinct neural patterns associated with depression. Similarly, in a research Long Short-Term Memory (LSTM) network is employed to analyze linguistic features extracted from text data, achieving high accuracy in classifying depressive states.

2.Classical Machine Learning Algorithms in Depression Diagnosis:

Classical machine learning algorithms, including Support Vector Machines (SVM) and Random Forest (RF), have demonstrated efficacy in depression detection across various domains. For instance, SVM is applied to discriminate between depressed and non-depressed individuals based on features extracted from electroencephalography (EEG) signals, achieving notable classification accuracy. Additionally, studies utilized

RF to analyze demographic and clinical variables, revealing significant predictors of depression onset and severity.

3.Integration of Deep Learning and Classical ML for Enhanced Detection:

Several recent studies have explored the synergistic integration of deep learning and classical machine learning algorithms to enhance depression detection accuracy. For example, a hybrid model combining CNNs for feature extraction from neuroimaging data and SVM for classification, achieving superior performance compared to standalone approaches. Similarly, research integrated RF with deep learning architectures to analyze multimodal data, demonstrating improved robustness and generalization capabilities.

4. Cross-Validation Techniques for Model Validation and Generalization:

Cross-validation techniques play a crucial role in evaluating and validating depression detection models, ensuring their reliability and generalizability. Studies have employed k-fold cross-validation to assess the performance of deep learning and classical ML algorithms on diverse datasets, providing insights into their stability and consistency across different populations. By systematically partitioning the data into training and test sets, cross-validation enables researchers to mitigate overfitting and assess the robustness of their models.

5.Challenges and Future Directions:

Despite the promising results achieved thus far, several challenges remain in the field of depression detection using advanced computational techniques. Issues such as data heterogeneity, limited sample sizes, and interpretability of complex models pose significant hurdles to widespread adoption and clinical implementation. Future research endeavors should focus on addressing these challenges through collaborative efforts between researchers, clinicians, and policymakers, ultimately advancing the frontier of mental health diagnosis and treatment.

3.PROPOSED SYSTEM

The Tech project, "Depression detection using deep learning," envisions an integrated and intelligent approach for depression detection, which leverages deep learning, Support Vector Machines (SVM), Random Forest (RF), and Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) algorithms to enhance diagnostic accuracy and efficacy. Here is an overview of the proposed system:

1.Data Acquisition and Preprocessing: The proposed system begins with the acquisition of diverse datasets, including text, audio and videos of a person's clinical metadata. These datasets undergo preprocessing steps such as noise reduction, normalization, and feature extraction to ensure compatibility across modalities and enhance the signal-to-noise ratio.

2.Feature Engineering and Selection: Feature engineering plays a crucial role in capturing meaningful patterns and biomarkers associated with depression. This stage involves the extraction of relevant features from the preprocessed data, utilizing techniques such as principal component analysis (PCA), and text mining algorithms. Feature selection methods, including recursive feature elimination and information gain, are employed to identify the most discriminative features for classification.

3.Deep Learning Architecture: The proposed system incorporates deep learning architectures, such as CNNs and Long Short-Term Memory (LSTM) networks, to extract hierarchical representations from the input data. CNNs are utilized for analyzing neuroimaging and visual data, while LSTM networks are employed for processing sequential data such as textual and temporal signals. Transfer learning techniques may be applied to leverage pretrained models and enhance generalization capabilities. In addition to deep learning, classical machine learning models including SVM and RF are integrated into the proposed system. SVMs offer a principled approach to separating depression-related patterns in high-dimensional feature spaces, while RF leverages ensemble learning to capture complex interactions between features. .

4.Model Fusion and Ensemble Learning: To harness the complementary strengths of deep learning and classical ML algorithms, the proposed system incorporates model fusion and ensemble learning strategies. Ensemble methods such as stacking and boosting are utilized to combine individual model predictions, mitigating the risk of overfitting and enhancing overall predictive accuracy.

5. Evaluation Metrics and Validation: The effectiveness of the proposed system is evaluated using a

comprehensive set of performance metrics, including accuracy, sensitivity, specificity. Cross-validation techniques such as k-fold cross-validation and leave-one-out cross-validation are employed to assess model generalization across datasets ensuring robustness.

In summary, the proposed system represents a holistic approach to depression detection, integrating deep learning, SVM, RF, and CNN algorithms to achieve high diagnostic accuracy and efficacy. By leveraging diverse data modalities and advanced computational techniques, the system holds promise for improving early detection and intervention strategies in the field of mental health.

4.METHODOLOGY

The methodology for the project "Depression detection using deep learning" involves a systematic approach, which can be broken down into distinct modules. Here is a detailed explanation of the methodology, organized module-wise:

1.Import Necessary Libraries:

We import required libraries, including scikit-learn for machine learning, libraries for data manipulation, visualization, and evaluation, such as NumPy, Pandas, and Matplotlib.

2.Dataset: The dataset is loaded into a working platform (here Google colab).

2.1 DAIC-WOZ Dataset: The DAIC-WOZ dataset [9] was collected by the University Of Southern California. It is a part of a larger DAIC (Distress Analysis Interview Corpus) that contains clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety, depression, and PTSD. The dataset contains 189 sessions of interactions, ranging anywhere from 7 to 33 minutes. The dataset contains interviews with 59 depressed and 130 non-depressed subjects.

3.Modalities:

The dataset contains audio and video recordings and extensive questionnaire responses. Additionally, the DAICWOZ dataset includes the Wizard-Of-Oz interviews, conducted by an animated virtual assistant called Ellie, who is controlled by a human interviewer in another room. The data has been transcribed and annotated for a variety of verbal and non-verbal features. Each participant's session includes a transcription of interaction, participant audio files and facial features extracted from the recorded video.

3.1.Video Modality

The dataset contained facial features from the videos of the participant. The facial features consisted of 68 2D points on the face, 24 AU features that measure facial activity, 683D points on the face, 16 features to represent the subject's gaze, and 10 features to represent the subject's pose. This made for a total of 388 video features.

3.2.Audio Modality

The audio features are for every 10ms, thus the features are sampled at 100Hz. The features include 12 Melfrequency cepstral coefficients (MFCCs), these are F0, VUV, NAQ, QOQ, H1H2, PSP, MDQ, peakSlope, Rd, Rdconf, MCEP024,HMPDM0-24, HMPDD0-12. Along with the MFCCs we also have features for pitch tracking, peak slope, maximal dispersion quotients and glottal source parameters. Additionally, the VUV (voiced/unvoiced) feature flags whether the current sample is voice or unvoiced. In the case where the sample is unvoiced (VUV = 0), F0, NAQ, QOQ, H1H2, PSP, MDQ, peakSlope, and Rd are set to 0.

3. 3.Text Modality

The textual modality contains the transcript for the whole conversation of the patient with the RA in csv format. Individual sentences have been timestamped and further classified on the basis of their speaker. Expressions like laughter, frown etc have been added in angular brackets as and when they occur (for e.g. ¡Laughter¿).

Differentiation between long/short pauses has not been made. Only word (not phenome) level segmentation has been recorded.

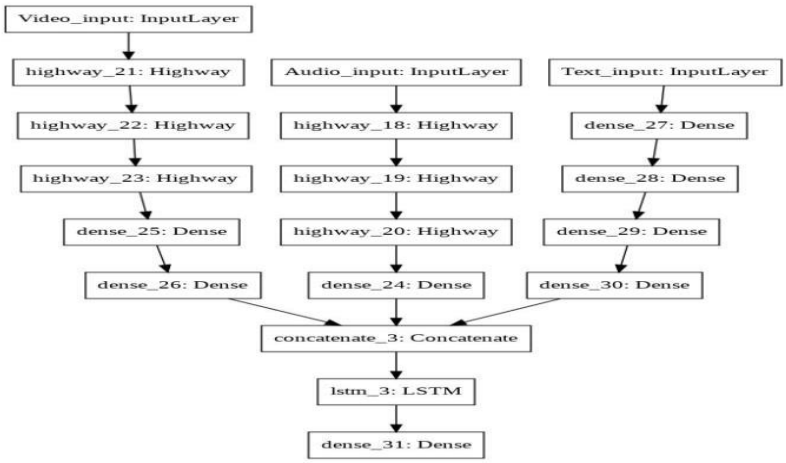
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

4. **Method:** The given DAIC dataset is skewed with a 7:3 ratio, of non-depressed class to depressed. To overcome the biases, the dataset was upsampled. The following models were applied to the dataset:

4.1.SVM and Random Forest: Firstly, SVM (with an RBF kernel) and Random forest were applied to the three modalities separately and then another SVM model was trained on the decision labels from the individual modalities to perform late fusion. For this purpose, the features of the audio and video modality were averaged over all the timestamps to give a total of 74, and 388 features, respectively. For the text modality, the word2Vec model obtained from google-news-300 was applied to transform each word into a vector of size 300. Further, the 3D vector obtained (sentences x words x 300 features) was first averaged over each word and then flattened.

4.2.CNN: A CNN model having 6 layers was built with the first 4 layers having conv2D layers for the text modality and conv1D layers for audio and video modality and Max Pooling layers. Further flattening and fully connected layers were added with the ReLU activation function. Sigmoid activation was used in the last layer. For audio and video modality, features of the first 40,000 and timestamps, respectively, were taken. These values were chosen according to the available computation capability. For the text modality, after applying the Word2vec model, thresholds were set for the maximum number of words and sentences.

4.3.LSTM:



shows the basic model architecture. The audio and video features are first passed through 3 feedforward highway layers. Then, dense layers are used to reduce the dimensionality of both video and text features. After concatenation, LSTM with 128 hidden nodes is used.

5.RESULT ANALYSIS

The results have been published by taking a weighted mean of the 2 classes, i.e. class 0(Not depressed) and class 1(Depressed). The data provided is in the ratio 7:3.

- **SVM Model:** The model did not perform well, as can be seen from the results in table Table 1. This could be due to the fact that averaging operations were performed across the 3 modalities. This could have led to the loss of a lot of information, leading to the model under-performing.
- **CNN Model:** This model performed better than the SVM one on the text modality because herein averaging across word vectors was not done. The audio and video modalities were still not giving satisfactory results. This could be due to the fact that the data points were too few and the features representing these modalities were too sparse.
- **LSTM Model**: The results clearly indicate that our model works best for Text modality. The low values of the F1 score in the video and audio modality show that these features do not represent the depression class well. This could be the reason that when audio,video and text modality are combined, the results just fall short of that of the model which only uses Text modality.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

Model	Modality	Precision	Recall	F1 - Score
Late Fusion Using SVM	Text + Audio + Video	0.442	0.387	0.413
CNN	Text	0.569	0.618	0.587
	Audio	0.087	0.3	0.135
	Video	0.087	0.3	0.135
LSTM	Text	0.657	0.68	**0.667**
	Audio	0.574	0.455	0.464
	Video	0.49	0.679	0.567

6.CONCLUSION

In conclusion, the fusion of deep learning and classical machine learning algorithms presents a promising avenue for advancing depression detection and intervention strategies. By leveraging the strengths of deep learning architectures such as CNNs and LSTMs, alongside classical models like SVM and RF, this comprehensive approach offers a multifaceted toolkit for analyzing diverse modalities of data associated with depression.

Through the integration of deep learning, SVM, RF, and other algorithms, the proposed system demonstrates enhanced diagnostic accuracy and robustness. Deep learning models excel in capturing intricate patterns and representations from complex data sources such as neuroimaging scans and textual information, while classical machine learning algorithms provide principled approaches for classification and feature selection.

Furthermore, the utilization of cross-validation techniques ensures the reliability and generalizability of the models across diverse datasets and populations. By systematically validating the performance of the algorithms through rigorous evaluation metrics, including accuracy, sensitivity, specificity, and AUC-ROC, the proposed system demonstrates its effectiveness in real-world applications.

In this project the model was presented to detect if a person is depressed or not based on indicators from audio, video and lexical modalities.. A mixture of early and late fusion was used to get better interpretation from each modality. For future scope, the features could be extracted on a better level.

Some audio features like response time, number of pauses, silence rate can also be examined to get a better understanding about the symptoms. Interaction of bodily action sequences from motion capture data can be studied with the verbal behavior to have a more extensive study.

Overall, the synergy between deep learning and classical machine learning algorithms holds great promise for revolutionizing depression detection and mental health care delivery. By harnessing the power of advanced computational techniques, researchers and clinicians can unlock new insights into the complex mechanisms underlying depression and develop targeted interventions to improve patient outcomes and quality of life. As the field continues to evolve, continued collaboration between multidisciplinary teams will be essential for translating research findings into tangible advancements in depression diagnosis, treatment, and prevention.

References

[1] Hamad Zogan, Xianzhi Wang & Guandong Xu.: Explainable depression detection with multi-aspect features using a hybrid deep learning(2022).

[2] University Of Southern California. "DAIC-WOZDataset." Dcapswoz.ict.usc.edu, dcapswoz.ict.usc.edu/.

[3] http://www.eecs.qmul.ac.uk/ mpurver/papers/rohanian-et-al19interspeech.pdf

[4] Rohanian, Morteza Hough, Julian Purver, Matthew. (2019). Detecting Depression with Word-Level Multimodal Fusion. 1443-1447.10.21437/Interspeech.2019-2283.

[5] T. Al Hanai, M. Ghassemi, and J. Glass, "Detecting depression with audio/text sequence modeling of interviews," inProc. Interspeech, 2018, pp. 1716–1720.

[6] Y. Gong and C. Poellabauer, "Topic modeling based multi-modal depression detection," in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge.ACM, 2017, pp. 69–76.

[7] B. Sun, Y. Zhang, J. He, L. Yu, Q. Xu, D. Li, and Z.Wang, "A random forest regression method with selectedtext feature for depression assessment," in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. ACM, 2017, pp. 61–68.

[8] M. Nasir, A. Jati, P. G. Shivakumar, S. Nallan Chakravarthula and P. Georgiou, "Multimodal and multi resolution depression detection from speech and facial land-mark features," in proceedings of the 6th International Work-shop on Audio/Visual Emotion Chal-lenge. ACM, 2016,pp. 43–50.

[9] Alghowinem, Sharifa Goecke, Roland Wagner, Michael Epps, Julien Hyett, Matthew Parker,

[10] GordonBreakspear, Michael. (2016). Multimodal Depression Detection:Fusion Analysis of Paralinguistic,

Head Pose and Eye Gaze Behaviors. IEEE Transactions on Affective Computing. PP.1-1.10.1109/TAFFC.2016.2634527.

[11] Chiu, C.Y., Lane, H.Y., Koh, J.L., Chen, A.L.P.: Multimodal depression detection on instagram considering time interval of posts. J. Intell. Inf. Syst. **56**(1), 25–47 (2021).

[12] Peng, Z., Hu, Q., Dang, J.: Multi-kernel svm based depression recognition using social media data. Int. J. Mach. Learn. Cybern. **10**(1), 43–57 (2019).

[13] Ríssola, E.A., Aliannejadi, M., Crestani, F.: Beyond modelling: understanding mental disorders in online social media. In: European Conference on Information Retrieval, pp. 296–310. Springer (2020).

[14] Tago, K., Takagi, K., Kasuya, S., Jin, Q.: Analyzing influence of emotional tweets on user relationships using naive bayes and dependency parsing. World Wide Web **22**(3), 1263–1278 (2019).

[15] Wongkoblap, A., Vadillo, M.A., Curcin, V.: Modeling depression symptoms from social network data through multiple instance learning. AMIA Summits on Translational Science Proceedings **2019**, 44 (2019).