Journal of Information Systems Engineering and Management
2025, 10(31s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Echoes of the Mind: A CNN Approach for Early Mental Health Prediction

Vanita G. Kshirsagar[1*], Ravindra S. Apare[2], Sheetal S. Patil[3], Kalyan D. Bamane[4], Arati K. Kale[5], Gajanan P Arsalwad[6], Sunil Kumar Yadav[7]

[1]Dr. D. Y. Patil Institute of Technology, Pune,*
[2]Trinity College of Engineering and Research, Pune, India
[3] MIT College of Management, MIT ADT University, Loni kalbhor, Pune
[4] D Y Patil College of Engineering Akurdi
[5] KJEI's Trinity College of Engineering and Research, Pune
[6] Trinity College of Engineering and Research, Pune
[7]Department of Computer Science and Engineering, Amity University, Rajasthan

ARTICLE INFO

Received: 29 Dec 2024

Revised: 12 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Introduction:

The depression often goes undiagnosed due to the absence of objective and accessible detection methodology. This paper focuses on developing an audio-based system that analyzes speech patterns, tone, and sentiment to predict early signs of depression, enabling timely intervention.

Objectives:

To design a machine learning model that detects depression using speech features such as pitch, tone, and rhythm. To improve early mental health diagnosis by leveraging audio-based sentiment analysis.

Methods:

Speech signals will be processed using feature extraction techniques like MFCCs (Mel-Frequency Cepstral Coefficients) and spectral analysis. Deep learning models, like LSTM or CNN, will classify speech patterns to identify human behavior.

Results:

The model accurately distinguishes between sad and neutral speech. Audio-based sentiment analysis demonstrates that it must be a useful tool for early-stage mental health assessments.

Conclusion:

Depression identification using speech is a non-invasive and scalable option for mental health screening. This method improves early diagnosis of depressive behavior and makes mental health monitoring more accessible and data-driven.

Keywords: Mental Health, MFCCs (Mel-Frequency Cepstral Coefficients) , CNN, LSTM,Deep Learning (DL).

INTRODUCTION

In the modern era, depression is one of the most common mental health issues prevailing in millions of lives around the world. The diagnosis can be done basically through the ways in which people describe their feelings, but it doesn't always give the whole and accurate picture. Through machine learning, we can find new approaches to analyze non-verbal signs such as speech patterns, text, and facial expressions. [4]. With the rise of online social networks, sentiment analysis has become increasingly important. Sentiment analysis, commonly called opinion mining, leverages techniques like natural language processing (NLP), text mining, and text analysis to uncover and interpret subjective insights from the content provided. [8]. Much of the research in sentiment analysis leverages e-text from the web, utilizing techniques such as machine learning, support vector machines, latent sentiment analysis, and text mining. [10] In speech analysis, noise from signals produced by speakers can hinder accurate assessment, making noise removal essential before analysis. Techniques like Notch Filter, Noise Gate, and Much of the research in sentiment analysis leverages text from the web, utilizing techniques such as machine learning, support vector machines, latent sentiment analysis, and text mining. [5] In speech analysis, noise from signals produced by speakers can hinder accurate assessment, making noise removal essential before analysis. Speech signals have unique advantages in affective computing. Compared to facial expressions, which can be altered by physical movements or obscured by glasses or beards, speech qualities are less affected by such factors. Because speech carries valuable information, automated systems that process a patient's speech signals can provide clinicians with intelligent feedback, assisting in clinical assessments. [6]

Noise gate plug-ins are used for this purpose. Speech signals have unique advantages in affective computing. Compared to facial expressions, which can be altered by physical movements or obscured by glasses or beards, speech qualities are less affected by such factors. Because speech carries valuable information, automated systems that process a patient's speech signals can provide clinicians with intelligent feedback, assisting in clinical assessments. Recognizing emotions in mental health patients offers insights that help deliver effective care. However, it is challenging for practitioners to manually recognize emotions. Depressed individuals behave differently from non-depressed individuals, a difference discernible in their speech recordings. They often avoid eye contact, engage less in verbal communication, speak tensely in short phrases, and feel uncomfortable in conversations. [7] Such systems mainly aim to improve diagnostic accuracy by incorporating supplementary data, like emotion detection through human voice analysis. Despite the benefits, social stigma associated with mental disorders often hinders early diagnosis and access to medical services. To address this, automated mental care services have been proposed, using emotional analysis to detect depression. Consistently tracking individuals' daily activities within their natural surroundings can enable the automated collection and analysis of data, leading to faster and more precise evaluations of depressive symptoms. [8] Enhancing assessment methods in this way could significantly reduce the widespread social and economic effects of depression. It not only identifies sentiment-related words but also detects relationships between words to accurately determine sentiment. Currently, various approaches are available for sentiment extraction. Depressed individuals are more inclined to tension, trouble, and sadness and are every now and again stressed and unengaged. People struggling with depression often find it difficult to focus on their tasks and interact with others and may even withdraw socially. Research indicates that many individuals experiencing depression are unaware that they are actually suffering from it. [11]

The objective of this research is to develop an automated system capable of analyzing a person's voice tone to detect signs of depression. By identifying non-verbal cues linked to depression, this system could assist doctors and mental health professionals in diagnosing depression more quickly and accurately. Early detection of mental health issues through this approach can help mitigate potential future complications. The paper contains seven comprehensive sections. Section 1 provides an overview of the goals of the study and the relevance of mental health in the age of digital interactions. Section 2 gives an insight into the technologies employed and the problems constraining the processes of identifying mental health aspects. Section 3 summarizes the studies conducted in a certain field and mentions what has been done and what problems still need to be solved. Section 4 describes the approach adopted by the authors of this paper, which incorporates the use of ML and NLP techniques to enhance the performance of the sentiment analysis. Section 5 describes the results of the experiment designed to demonstrate the effectiveness of the techniques discussed above. Section 6 draws the main conclusions of the given research and provides some guidelines for further work in the respective area; the issue of AI and its significance with respect to mental health studies, as well as how this sphere deserves further investigation, is discussed. Section 7 gives the references of all the papers taken for review.

OVERVIEW

The most widely used tools and methods we will use to identify depression by voice. How vocal characteristics may be used differently to possibly classify emotional states.

2.1 Preliminary Information

This paper introduces the basic features of audio use in the diagnosis of mental disorders. The use of techniques that examine aspects of the human voice in terms of pitch, tone, and rhythm, among others, to find the features that may associate with depression- or anxiety-based emotions. These basics have been refined and are used later in this paper.

2.2 Acoustic Feature Extraction

Extract audio into several parts, or features, that may give additional information about a person's voice. Main features are:

1) MFCCs (Mel-Frequency Cepstral Coefficients): These are widely known features that can capture the pitch and the tone of one's voice.

2) Chroma Feature: These will help in identifying the variation in pitch, which can indicate if he sounds low or sad.

3) Energy and Spectral Features: These include the energy that is used as well as the level of clarity in the voice of a certain person, which can be quite useful in finding emotional tones. For that purpose, use libraries such as Librosa and OpenSMILE.

2.3 Machine Learning Model

To categorize diverse emotions by training models on various voice features. Common methods include SVMs,**random forest,** and gradient boosting. SVMs are great for voice classification, such as depressed or not depressed. Random forests and gradient boosting are excellent at identifying patterns in audio features and are great at identifying a wide range of emotions. Logistic Regression Much simpler model, designed mainly for just a simple classification of voice features. These models learn from training data, helping them find patterns in new, unseen audio, which may indicate the existence of depression.

2.4 Audio Deep Learning

Deep learning can spot much more abstract patterns in voice data, particularly where such patterns evolve over time. Useful deep learning models include RNNs and LSTMs: These are the models that look into voice over time. This might help spot changes in tone or pauses that point out depression. These models allow for minute analysis over vocal patterns. This would be the major route to detecting slight changes in the voice as depression indicators.

2.5 Convolutional Neural Networks (CNNs) for Spectrogram Analysis

In fact, CNNs work quite well with analyzing audio as spectrograms, which indicate frequencies of sounds, such as this:

1) Input Spectrogram: Audio converted into image files, where pitch and tone over time are illustrated. The CNN scans those images to look for patterns that match with emotion.

2) Pattern Recognition: The CNNs detect the patterns in the above spectrogram that help make the speech pattern analysis about low or sad emotions.

CNNs allow for ease in the identification of particular patterns in audios that would eventually represent trends regarding depression.

2.6 Audio Transformers

Audio Transformers: These abstracts can be treated as the emotion detector that learns emotions correctly and understands low-level voices in speech, which further reveals mental conditions. Transformers significantly contribute to developing adaptable, powerful, yet flexible models across varied languages, accents, and speech.

2.3 Machine Learning Model

To categorize diverse emotions by training models on various voice features. Common methods include SVMs, random forests, and gradient boosting. SVMs are great for voice classification, such as depressed or not depressed. Random forests and gradient boosting are excellent at identifying patterns in audio features and are great at identifying a wide range of emotions. Logistic Regression Much simpler model, designed mainly for just a simple classification of voice features. These models learn from training data, helping them find patterns in new, unseen audio, which may indicate the existence of depression.

2.4 Audio Deep Learning

Deep learning can spot much more abstract patterns in voice data, particularly where such patterns evolve over time. Useful deep learning models include RNNs and LSTMs: These are the models that look into voice over time. This might help spot changes in tone or pauses that point out depression. These models allow for minute analysis over vocal patterns. This would be the major route to detecting slight changes in the voice as depression indicators.

2.5 Convolutional Neural Networks (CNNs) for Spectrogram Analysis

In fact, CNNs work quite well with analyzing audio as spectrograms, which indicate frequencies of sounds, such as this:

1) Input Spectrogram: Audio converted into image files, where pitch and tone over time are illustrated. The CNN scans those images to look for patterns that match with emotion.

2) Pattern Recognition: The CNNs detect the patterns in the above spectrogram that help make the speech pattern analysis about low or sad emotions.

CNNs allow for ease in the identification of particular patterns in audios that would eventually represent trends regarding depression.

2.6 Audio Transformers

Audio Transformers: These abstracts can be treated as the emotion detector that learns emotions correctly and understands low-level voices in speech, which further reveals mental conditions. Transformers significantly contribute to developing adaptable, powerful, yet flexible models across varied languages, accents, and speech.

LITERATURE SURVEY

3.1 Methodology for systematic literature review

Every keyword was needed to do article searches that were relevant to the poll. Finding pertinent publications was then made possible by searches conducted in a number of electronic databases, including Google Scholar, ACM, IEEE, ScienceDirect, and Springer.

Following a year and keyword filter, the chosen texts were categorized as research papers and review articles. examined the material in greater detail by analyzing the sentences as well as gathering data from the Microsoft Offi ce documents.

3.2 Findings of the Present Study

Lexicons are the lists of words or phrases that express sentimental values. As a result, they are used to interpret feelings or viewpoints that may be expressed in the text. Kostadin Mishevm, Ana Gjorgjevikj, and their group effectively employed transformers in conjunction with lexicons [12] to improve sentiment analysis accuracy. The models, like GPT, BERT, and other transformer models, improve the arrangement and meaning of the material by understanding the context of a text, going beyond lexicons. Text mining [13, 14] has been useful in the area of mental health studies for identifying emotional indicators such as stress or worry, mostly from social media platforms. Using these word sequences and advanced natural language processing approaches, Haruna Isah and colleagues [15]. Ordinal regression is a good statistical tool for ordered categories with an increasing level of severity but no exact interval between them. Shihab Elbagir and Jing Yang [23] used this strategy to improve sentiment analysis accuracy. Meanwhile, clustering [16, 17] is an unsupervised learning technique used to classify things into groups of similar objects in order to extract patterns. For example, Shreya Ahuja, Gaurav Dubey, and Hima Suresh [17] employed clustering approaches to identify sentence polarity and extract sentiment from their work. Homogeneous ensemble classifiers [18] are an excellent method for boosting model performance since they combine numerous models of the same type, such as decision trees, to reduce overfitting and improve accuracy. To increase the model's resilience, these methods include bagging and boosting. Due to their ability to observe the complete context of words in phrases, models such as BERT are helpful for contextual analysis tasks. BERT is a more accurate model than Naive Bayes, which is nevertheless useful for simple text classification tasks despite its simplicity, according to researchers like Tianyi Wang and Ke Lu [19]. According to Sandy Kurniawan et al. [20], many researchers have combined machine learning (ML) and deep learning (DL) techniques in hybrid approaches to improve sentiment analysis outcomes as an alternative to employing these models alone. Yogesh Chandra and Antoreep Jana [21] integrated the models and employed hybrid approaches to improve the outcomes.

DL, the most important machine learning technology, uses multi-layered neural networks to allow the model to handle raw data, such as text, audio, or images, on more complicated tasks. To improve accuracy, Mehmet Umut

Salur and Ilhan Aydin [22] used hybrid deep learning (DL) models, while Hasibe Busra Dogru et al. [23] looked at the Doc2Vec model to improve the text representation accuracy. The best way is to combine the LSTM with dense layers for emotion recognition, which has been the subject of numerous studies. With an accuracy score of 99% versus 10 epochs using simply LSTM [24], a study by Dr. C. S. N. Murthy et al. [25] was more effective. The capacity to generate convolutional filter layers, which can extract complex characteristics from the image, has also made convolutional neural networks [26-32] seem like a good fit for the task. This CNN structure guarantees effective feature identification, from edges to intricate textures, as it can be applied to images of any complexity and yet identify them. More head mechanisms were incorporated into CNN models by Yue Feng [30] in order to increase accuracy. Jin Wang [20] combined CNN with LSTM in modeling a better model. Zabit Hameed and Begonya Garcia-Zapirain [33] found that BiLSTM performs well at a high epoch count of 100. Furthermore, Guixian Xu et al. [41] achieved high accuracy using BiLSTM and the ReLU activation function over 200 epochs. These models show that complex neural networks are well organized for sentiment analysis and text classification. The research highlights the methods such as BERT, CNN-LSTM hybrids, and BiLSTM used to progress accuracy by leveraging feature difficulty and background. Several studies have shown that amending machine learning or deep learning methods, whether with new layers or in combination with present models, can greatly enhance precision in sentiment analysis tasks. The current literature has thus served as a foundation for further improving designs based on model complexity. Numerous studies have presented how changes to machine learning or deep learning methodology, whether adjusted with new layers or used in tandem with current models, can significantly improve precision in sentiment analysis tasks. This existing literature has thus served as a foundation for further refining designs based on the model complexity of difficulties in text mining and emotion tasks [35-40].

METHODOLOGY

The proposed approach focuses on identifying and validating speech signals to detect depression and other mental health disorders through speech signal processing using deep learning (DL) and convolutional neural networks (CNN), which is represented in Fig. 1. This method utilizes extensive datasets comprising numerous speech samples, as outlined in the architecture model. The primary goal of this technique is to establish a correlation with trained datasets to extract, evaluate, and classify speech samples as needed. These speech signals are interconnected and exhibit a high level of distinction in assessing and validating emotional states and mental stability.

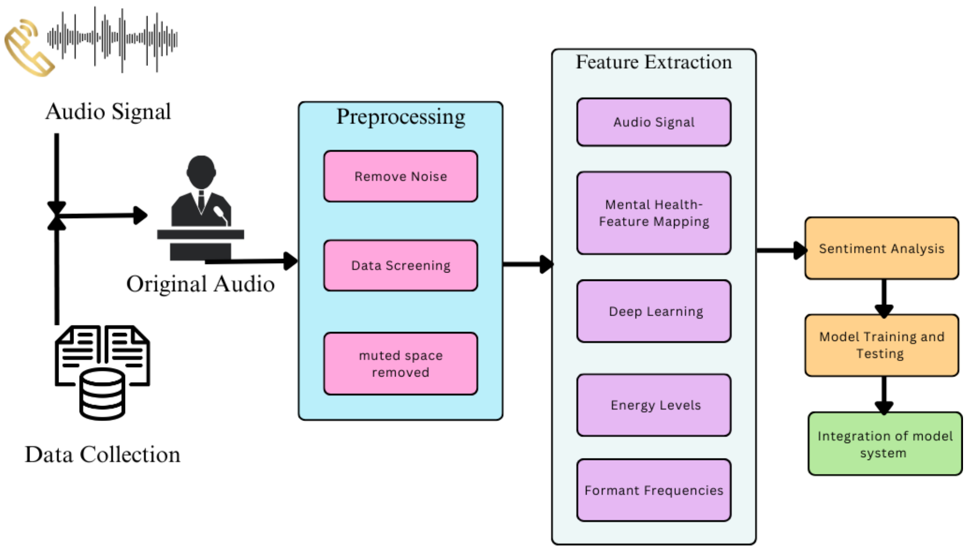

Fig. 1 Mental health disorders through speech signal processing using deep learning (DL)

1. Data Collection

The participants' voice samples have been gathered in both controlled and natural settings. They have been inquired to speak about their everyday life happenings, thoughts, and emotions. These samples have been recorded using good-quality microphones so that clarity and accuracy can be maintained and stored in a database that was easily accessible for further research. Diversification has been noticed in the sample since participants belong to different age groups, genders, and backgrounds.

2. Preprocessing of Data

The collected raw voice data was thus pre-processed to remove noise or irrelevant sounds that could interfere with the analysis. The usual audio processing techniques for doing this are noise reduction, normalization, and silence removal. The isolation of relevant vocal segments is performed while trying to maintain the quality of the voice signal as high as possible.

3. Feature Extraction

Extractive feature extraction from the processed voice data: we were to extract features that had been associated with emotional states, such as: Pitch Changes in pitch might mirror intensity changes in emotion. Speech rate Patterns that are slow and monotonous often correlate with depression.

Energy levels: Reduced energy in voice can be symptomatic of depression.

Formant frequencies: These could manifest alterations in the quality of speech and resonance, which commonly suffer in depression. We also used features that were automatically extracted from audio samples using tools such as LibROSA.

4. Sentiment Analysis

After feature extraction, the voice data were analyzed in terms of sentiment, wherein all those speech samples were checked to determine the emotional tone of the speech. Classifications of the voice recordings into neutral, happy, sad, or anxious were conducted by using various models of machine learning, including CNN and RNN. The models were then trained on annotated datasets where each sample of voice was pre-labeled according to the speaker's emotional state.

5. Model Training and Evaluation

We used the features extracted and sentiment data in a machine learning model designed with that requirement in mind: identifying markers of depression. We trained a deep learning model, for example, such as CNN or LSTMs, using labeled data of participants diagnosed at different degrees of depression. To evaluate the model's ability to determine depressive states solely from voice data, a range of standard metrics was employed, including accuracy, precision, recall, and the F1 score.

6. Integration of Model System

Finally, we integrated results from the voice analysis with other modalities, such as facial expression analysis and text-based sentiment analysis, into one comprehensive model that detects depression. We then combined the audio sentiment with these other inputs to enhance the overall system accuracy and reliability in detecting depression. This method of voice data analysis also focuses on capturing subtle vocal indicators of distress in this manner of somber approach to detecting depression. The objective evaluation with machine learning. In these sections, we present the result of the experimentation of the depression detection model based on audio-based emotion analysis. The public datasets, as well as a private dataset of voice recordings from depressed patients, have been used for the identification of depressive vocal characteristics and differentiation between depressed and non-depressed emotional states.

EXPERIMENTAL RESULTS

A. Dataset Overview

Audio Corpus The audio corpus contains recordings of people with other emotional expressions, with particular emphasis placed on depressive signs. It is divided into two control group audio samples of people who are not depressed. In this case, the tone is largely neutral or even positive. Depressive group recordings of depressed people. The speech rate is slower, less energetic, and vocal manifestations of depression/indifference are included in the corpus.

B. Model Training

Audio analysis used machine learning and deep learning techniques applied during model training.

Feature Extraction: Mel-Frequency Cepstral Coefficients MFCCs that pick those important features of the sound talking about the pitch, tone, and rhythm of the audio. We also included some of the voice characteristics like pitch changes and pauses to capture or better understand emotional tones. The RNNs and LSTM layers helped in the pattern extraction of emotions over time. Essentially, the RNNs or the LSTM layers define the sequences as well as the timing involved with speech. For training purposes, we used gradient-based training of which some data have transformations; thereby, increasing variety in terms of training.

C. Evaluation Metrics

The model was evaluated using the following metrics:

1) Accuracy: This measures how often the model correctly classified emotional states as either depressed or not depressed.

2)Precision: Number of those it diagnosed as depressive, which were actually depressive.

3) Recall: Number of actual depressive that it captured.

4) F1-score: trade-off between precision and recall, gives the overall correctness of model.

On test set, the table 1. Evaluation Metric represents results are follows:

Table 1. Evaluation Metric

Metric	Value
Accuracy	87.5%
Precision	86.1%
Recall	90.4%
F1-Score	88.2%

These results indicate that the model is good in classification in both states of depression and not, with a good accuracy in classifying depressive vocal tones.

D. Results of Cross Validation

We ran 5-fold cross-validation to check if the models are consistent. The model was consistent throughout folds, with an average accuracy of checking stability in determining depressive symptoms using audio samples at 86.3%.

E. Comparison with Baseline Models

We evaluate the performance of the RNN-LSTM model in comparison to traditional machine learning models.

1) SVM: With accuracy of 79.2%

2) RF: Achieved accuracy of 82.5%

3) LR. Accuracy: 77.8%

Comparative Analysis:

The outperform all the baseline models with developed RNN-LSTM model that obtained the highest accuracy, precision, recall, and F1-score values. This validates that deep learning-based methods are better suited for the task of emotional cue detection from audio data relevant to depression.

RESULTS AND DISCUSSION

Depression detection based on social media sentiment analysis is promising, with a suitable algorithm that analyzes data on Twitter to classify emotions and score feelings such as happiness and sadness, hence validating its utility in identifying depression patterns and acting as a worthwhile tool for mental health evaluation [4]. Acoustic features have also been employed quite efficiently for the identification of depression. CNN-based spectrogram analysis has resulted in achieving an accuracy of 82% on training data and 78% on test data. It avoided the analysis of speech content and, therefore, guaranteed privacy. It can also be incorporated seamlessly with smart assistants. This makes it a reliable technique for the diagnosis of mental health conditions [41]. Multimodal analysis further enhances the detection framework by including voice and text data. A model that used BERT-CNN for text processing and CNN-BiLSTM with attention mechanisms for voice showed improved accuracy by mitigating biases inherent in single-mode analysis, demonstrating scalability and stability for broader applications beyond clinical datasets [9]. Another text-voice multimodal method using the DAIC-WOZ dataset achieved an F1 score of 0.8 for the text and 0.76 for voice quality and pointed out that an early non-invasive diagnosis can be made by obtaining complementary features from both modalities [6].

Further enhancement was brought in by hybrid models that combine textual as well as audio along with CNN and Bi-LSTM that reach up to an accuracy of 98% for the model containing audio CNN and 92% for the text CNN model. Still, though improving predictive strength, the hybrid Bi-LSTM model was also longer in time for being trained, and that brings the use of multimodalities for the development of a better mental health predictor [42]. The integrated text, audio, and video modalities in emotion classification achieved 70% accuracy using weighted fusion techniques, though challenges included scarcity of datasets and processing complexity, and such improvements are needed for further developments in sentiment classification systems [7]. Hybrid speech recognition systems have turned out to be reliable, cost-effective, and, most importantly, applicable in healthcare contexts. For example, with the system using DTW and SVM, 97% accuracy was reported for systems compared to traditional systems with SVM that reported merely up to 79%. The system improved user authentication in handling appliances but could not successfully recognize unclear or affected voice conditions and hence needs fine-tuning to address these noise and diverse user scenarios [43]. Speech recognition has significantly increased productivity within healthcare with up to an 81% reduction in report turnaround times. Despite these, challenges have existed, such as hosting accented voices and handling standardized terminologies, and among the highlighted items are macros and templates, which help in streamlining workflow improvements [44]. Speech feature analysis on depression has been proven suitable with acoustic markers such as pitch and jitter. Making use of a large corpus of depressed, high-risk, and healthy subjects, the paper established in vivo a relationship between speech characteristics and depressive states, thus enabling the establishment of speech-based diagnostic tools. Yet feature selection has to be advanced to better enhance the quality of predictions to ensure its reliability and effectiveness in diagnosing mental illness [45].

CONCLUSION AND FUTURE SCOPE

The combination of speech and facial recognition might raise estimate correctness, while sentiment assessment using evidence from Twitter has shown great ability while doing an analysis of depression. Mental health monitoring becomes more manageable and adjustable with real-time monitoring through smartphone apps. The increased dataset improves the classification model, which will make it more healthy and practical. By combining text, audio, and video data, multimodal approaches can further expand diagnostic tools, which makes them more useful and accessible to a variety of consumers. The computational problems need to be resolved, and improving model scalability will contribute to the development of more thorough and reliable mental health assessment instruments.

FUTURESCOPE

Succeeding research has to focus on refining audio-based mental health identification through the integration of multilingual and culturally sensitive models. Real-time analysis united with social media and mental health apps will allow for quick involvements and better assistance for people experiencing difficulties.

REFRENCES

[1] Atul Tyagi, Nidhi Chandra, (2015).A Proposed Approach with Analysis of Speech Signals for Sentiment Detection

[2] Chiara Zucco, Barbara Calabrese, Mario Cannataro, (2017). Sentiment Analysis and Affective Computing for Depression Monitoring.

[3] Himani Negi, Tanish Bhola, Manu S Pillai, Deepika Kumar, (2018).A Novel Approach for Depression Detection using Audio Sentiment Analysis.

[4] Akriti Sood, Madhurima Hooda, Saru Dhir, Madhulika Bhatia, (2018). An Initiative to Identify Depression using Sentiment Analysis: A Machine Learning Approach.

[5] Zeenat Tariq, Sayed Khushal Shah, Yugyung Lee,(2019).Speech Emotion Detection using IoT based Deep Learning for Health Care.

[6] Vanita Ganesh Kshirsagar, Nishant Pachpor, Mrudul Arkadi, Nilesh Ghavate, Anita Mahajan, Ravindra Sadashivrao Apare, Sunil Kumar Yadav. *Early Prediction of Depression by using Text and Deep Learning*.

Journal of Information Systems Engineering & Management. 2025;10(15s):357-365. DOI: 10.52783/jisem.v10i15s.2464.
[7] Dr. Ashwini Rao, Akriti Ahuja, Shyam Kansara,Vrunda Patel,(2021).Sentiment Analysis on User-generated Video, Audio and Text.
[8] Pansy Nandwani, Rupali Verma, (2021). A Review on Sentiment Analysis and Emotion Detection from Text.
[9] Junhee Park, Nammee Moon, (2022).Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis.
[10] Vanita G. Kshirsagar, Sunil Yadav, Nikhil Karande. Feature Fusion and Early Prediction of Mental Health Using Hybrid Squeeze-MobileNet. International Advanced Computing Conference (IACC); 2023:417–426. DOI: 10.1007/978-3-031-56700-1_33.
[11] Ramanarayanan, V. (2024). Multimodal Technologies for Remote Assessment of Neurological and Mental Health.
[12] Kostadin Mishev, Ana Gjorgjevikj, Irena Vodenska, Lubomir T. Chitkushev, Dimitar Trajanov, 2020. Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. IEEE.
[13] Jalel Akaichi, Zeineb Dhouioui, Maria José López-Huertas Pérez, 2013. Text Mining Facebook Status Updates for Sentiment Classification. IEEE.
[14] Haruna Isah, Paul Trundle, Daniel Neagu, 2014. Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis. IEEE.
[15] Shihab Elbagir, Jing Yang, 2019. Twitter Sentiment Analysis based on ordinal Regression. IEEE.
[16] Shreya Ahuja, Gaurav Dubey, 2017. Clustering and Sentiment Analysis on Twitter Data. TEL-NET.
[17] Hima Suresh, Dr. Gladston Raj.S, 2016. An Unsupervised Fuzzy Clustering Method for Twitter Sentiment Analysis. IEEE.
[18] Murni, Tri Handhika, A. Fahrurozi, Ilmiyati Sari, Dewi P. Lestari, Revaldo Ilfestra Metzi Zen, 2019. Hybird Method for Sentiment Analysis Using Homogeneous Ensemble Classifier. IC2IE.
[19] Tianyi Wang, Ke Lu, Kam Pui Chow, Qing Zhu, 2020. COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model. IEEE.
[20] Sandy Kurniawan, Retno Kusumaningrum, Melnyi Ehonia Timu, 2018. Hierarchical Sentence Sentiment Analysis Of Hotel Reviews Using The Naïve Bayes Classifier. IEEE.
[21] Huyen Trang Phan, Van Cuong Tran, Ngoc Thanh Nguyen, Dosam Hwang, 2020. Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Enesemblr Model. IEEE.
[22] Mehmet Umut Salur, Ilhan Aydin, 2020. A Novel Hybrid Deep Learning Model for Sentiment Classification. IEEE.
[23] Hasibe Busra Dogru, Sahra Tilki, Akhtar Jamil, Alaa Ali Hameed, 2021. Deep Learning-Based Classification of News Texts Using Doc2Vec Model. IEEE.
[24] Kshirsagar, V.G., Yadav, S.K., Karande, N., Patil, P. Early prediction of mental health using SqueezeR_MobileNet, International Journal of Ad Hoc and Ubiquitous Computing, 2024, 47(3), pp. 158–175
[25] Dr. C. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, 2020. Text based Sentiment Analysis using LSTM. IJERT.
[26] Jin Wang, Liang-Chih Yu, K. Robert Lai, Xuejie Zhang, 2019. Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis. IEEE.
[27] Rongchao Yin, Peng Li, Bin Wang, 2017. Sentiment Lexical-Augmented Convolutional Neural Networks for Sentiment Analysis. IEEE.
[28] Hyunwoo Yu, Eunsu Lee, Suk-Bok Lee, 2016. SymBiosis: Anti-Censorship and Anonymous Web-Browsing Ecosystem. IEEE.
[29] Xi Ouyang, Pan Zhou, Cheng Hua Li, Lijun Liu, 2015. Sentiment Analysis Using Convolutional Neural Network. IEEE.
[30] Yue Feng, Yan Cheng, 2021. Short Text Sentiment Analysis Based on Multi-Channel CNN With Multi-Head Attention Mechanism. IEEE.
[31] Jonatas Wehrmann, William Becker, Henry E. L. Cagnini, Rodrigo C. Barros, 2017. A Character-based Convolutional Neural Network for Language-Agnostic Twitter Sentiment Analysis. IEEE.
[32] Yazhi Gao, Wenge Rong, Yikang Shen, Zhang Xiong, 2016. Convolutional Neural Network Based Sentiment Analysis using Adaboost Combination. IEEE.

[33] Zabit Hameed, Begonya Garcia-Zapirain, 2020. Sentiment Classification Using a Single-Layered BiLSTM Model. IEEE.
[34] Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, Xu Wu, 2019. Sentiment Analysis of Comment Texts based on BiLSTM. IEEE.
[35] Zhao Jianqiang, Gui Xiaolin, 2018. Deep Convolution Neural Networks for Twitter Sentiment Analysis. IEEE.
[36] Vipin Kumar, Basant Subba, 2020. A TfidfVectorizer and SVM based sentiment analysis framework for text data corpus. IEEE.
[37] Mondher Boiazizi, Tomoaki Ohtsuki, 2017. A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter. IEEE.
[38] Gamgarn Somprasertsri, Pattarachai Lalitrojwong, 2010. Extracting Product Features and Opinions from Product Reviews Using Dependency Analysis. IEEE.
[39] V. G. Kshirsagar et al., Fun Soundify: Music Generation powered by AI, 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 2024, pp. 1-6, doi: 10.1109/ICCUBEA61740.2024.10774982.
[40] V. G. Kshirsagar et al., Generative AI Powered Forensic Device, 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 2024, pp. 1-6, doi: 10.1109/ICCUBEA61740.2024.10774908.
[41] Aradhana, M. P., Chander, S., Krishna, B. (2020)."Diagnosing Clinical Depression from Voice: Using Signal Processing and Neural Network Algorithms to Build a Mental Wellness Monitor."
[42] Vandana, Nikhil Marriwala, Deepti Chaudhary(2023)"A hybrid model for depression detection using deep learning."
[43] Johnson, M., Lapkin, S., Long, V., Sanchez, P. (2014). A Systematic Review of Speech Recognition Technology in Healthcare.
[44] Ismail, A., Abdlerazek, S., El-Henawy, I. M. (2020). Development of Smart Healthcare System Based on Speech.
[45] Liu, Z., Hu, B., Yan, L. (2015). Detection of Depression in Speech.