Depressive Sentiment Prediction Through Audio Data Using Convolution Neural Network

Pragya Vishwakarma[1], Prof. Priyanka Saxena[2]
[2] Prof.
[1, 2] Takshshila Institute of Engineering & Technology, Jabalpur, M.P.

Abstract- *Recent advances in artificial intelligence applications show a growing interest in the development of automated systems for detecting depression and diagnosing social interventions. In particular, there has been active research with the diagnosis of depression based on various methods in deep learning, which can use various data by combining different data. This thesis presents a audio-stream-based diagnostic model for depression, including convolutional neural network (CNN), using spoken language. One-dimensional features of the music were extracted using frequency Cepstral coefficients and gamma tone Cepstral coefficients, respectively. Extracted features are used for cnn based on the transformation model.*

Finally, according to the proposed Audio-stream model, a diagnosis of depression is made using a set of softmax values derived from the CNN learning models. Using the proposed design, tests were performed using the depression database and the depression research collection wizard from other databases. Test results show efficiency of the proposed model. This shows that the proposed model is good for diagnosing depression.

Keywords- Depressive Sentiment, Speech Data, Wave plot, Frequency plot, CNN, Depressive, non-depressive.

I. INTRODUCTION

The current era, particularly the previous two decades, might be referred to as the ‘‘age of big data,’’ in which digital data is becoming increasingly important in various disciplines, including research, technology, society, and healthcare [1]. These datasets provide significant obstacles to the computing resources and analytics framework by making the whole study challenging for quickly extracting valuable information. Therefore, creating an effective big data analytics framework is a crucial research issue to overcome these sorts of problems. There are intensively used a number of research works in creating a big data analytics framework, such as Apache Spark [2], Apache Hadoop [3], Apache Storm [4], and Apache Kafka [5], to solve healthcare problems.
Mental health diseases are health problems that deeply affect the lives of individuals and must be treated with care. If the psychological disorders that appear in the human, which includes the depression, are not detected in time, they will have negative effects on the body. If this is the case, it is of great importance for the mental health of society.

Machine learning has the potential to inform illness models, the discovery and development of innovative medicines that can affect disease, as well as prevention methods in psychiatry [6]. Machine learning algorithms used to make sense of data on the big data analytics platform reveal results for the solution. In the healthcare field, it promises to inform the discovery and development of new disease-healing treatments. Techniques used in the field of recognizing brain and mental disorders have created incomplete or erroneous representations, and the results have not been sufficient. It is important for public health that psychological disorders benefit from the developments to be obtained by creating big data analytics. It can be observed that psychological disorders, which are health problems that deeply affect the lives of individuals and need to be treated carefully, are more likely to occur in some periods of life. To mitigate the damage of this era, researchers and psychiatrists have an unprecedented chance to make use of intricate patterns in the brain, behavior, and disease utilizing machine learning techniques [7]. By using these analysis techniques, it is aimed to detect important psychiatric disorders early, identify related factors and develop preventive measures.

Depression symptoms can be reflected in various human activities and behaviors and different degrees [8]. One of the sources, which can help identify depression symptoms in individuals, is the use of language [9]. Cognitive and linguistic studies have numerously shown that people with depression use language features differently [9]. For example, they tend to use more first-person singular pronouns (I, me, or we) and more negatively valenced words [10]. Online social content is one source for automatic mental disorder detection as it is one of the platforms through which users communicate. In recent years, social networking platforms have been widely applied to study users’ behavior and have inspired various researchers to introduce new forms of health care solutions [11], [12]. Furthermore, the stigma surrounding depression can make individuals less willing to seek professional assistance, and they turn to less traditional sources such as social media. Social media can be an essential source of information about individuals' opinions and feelings in the study of depression [13]. More specifically, research has addressed depression detection at various levels of granularity and approached it from different standpoints. Several social network sites (SNSs), such as Reddit, Twitter, Facebook, and Weibo, have been utilized for research about depression and other mental state disorders, such as postpartum depression [14] and posttraumatic stress disorder (PTSD) [15].

Prior approaches to misery discovery have essentially taken a bottom-up approach to memorize and apply profound learning (dl) and machine learning (ml) strategies. While such subsymbolic artificial intelligence (AI) methods can provide valuable insights about word frequencies and statistical correlations, they are not sufficient to analyze narrative and understanding of dialog systems in sentiment analysis [16]. Although there has been advancement with natural language processing (NLP) methods using DL methods, the predictive power of such approaches is limited mainly because DL methods learn better from large sets of data. Besides, communication entails a broader range of contributors, including understanding the world, social norms, and cultural awareness. To address these challenges, recently, research in depression detection has taken top-down approaches to learn by applying symbolic AI methods such as logical reasoning. In specific, the cross breed combination of subsymbolic approaches with typical strategies has been appeared to actuate more important designs in characteristic dialect writings [16]. Hence, it is vital to integrate symbolic approaches to learning with subsymbolic approaches in tackling the task of automated depression detection. In addition to hybrid methods, another set of approaches that yields high accuracy are ensemble methods in which several learning methods are combined [17]. Ensemble methods have frequently achieved high performance in solving various predictive problem areas [17].

1.2 Early depression detection

Sadness could be a common and tall rate of mental clutters. According to the World Health Organization (WHO) research on depression [18], there are about 350 million people with depression worldwide, and the incidence is increasing year by year. According to epidemiological studies, the lifetime prevalence of depression was 6.9% [19]. The main symptoms of depression are long-term low mood, loss of interest and pleasure; however, severe depression patients may self-harm and commit suicide. Depression not only brings grave harm to the life of patients but also brings a burden to society. Early detection of depression is beneficial in reducing the loss of individuals and society [20]. Therefore, early detection and intervention of depression are essential.

Currently, the typical clinical diagnosis methods of depression are mainly based on patients' self-evaluation and psychiatrist' clinical diagnoses. At present, the commonly used clinical diagnostic tools mainly include the Diagnostic and statistical manual of mental disorders (DSM-5) [21], Hamilton Depression Scale (HAMD) [22], Beck Depression inventory (BDI) [23], and Patient Health Questionaire - 9 items (PHQ-9) [24].

Figure 1.1 below shows some of the approaches for depression detection.

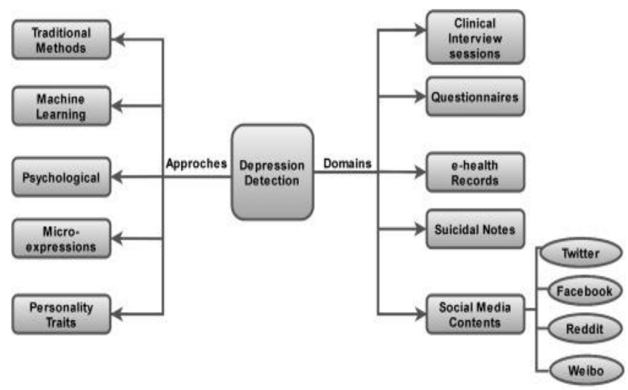

Figure 1.1: Depression Detection models.

II. RELATED WORK

In recent years, there has been active research on development of objective AI instruments for automatic depression diagnosis by analyzing physiological indicators such as facial expressions, audio features, EEG, and text data. There has been increasing interest in multi-modal analysis since the analysis enables acquisition of comprehensive and robust features, leading to a shift of research from single-mode to multi-mode analysis for depression diagnosis. Since depression is a complex and multifaceted disorder, a multi-modal approach for depression diagnosis and consideration of multiple modalities of information are essential in developing automatic depression diagnosis. Some of the existing literature on multi-modal approaches developed for depression diagnosis are outlined as follows.

Ay et al. [25] proposed a deep hybrid model developed using Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM) architectures to detect depression using EEG signals. In deep models, the temporal properties of signals are learned by the CNN layer and the sequence learning process is provided by the LSTM layer. In this work, they used EEG signals from the left and right hemispheres of the brain. They provided 99.12% and 97.66% classification accuracy for right and left hemisphere EEG signals, respectively. Therefore, it can be concluded that the developed CNN-LSTM model is accurate and fast in detecting depression using EEG signals.

Zhu et al. [26] proposed a method to distinguish between mild depression and normal controls using two feature fusion strategies (feature fusion and hidden layer fusion) for the fusion of EEG and EM signals based on the multimodal denoising autoencoder. They used the electroencephalography(EEG)-eye movement(EM) synchronous acquisition network to verify that the simultaneous recorded EEG and EM data during the experiment were synchronized with millisecond precision, which is the basis for meaningful analysis of EEG and EM data. Also, extracted 14 features (12 nonlinear and 2 linear features) for each band, Delta (0–0.2 Hz), Seta (0.2–0.4 Hz), Alpha (0.4–0.6 Hz), Beta (0.6–0.8 Hz), and Gamma (0.8–1 Hz). Finally, in relation to the multimodal autoencoder, their studied two feature fusion methods (Feature Fusion and Hidden Layer Fusion) to achieve fusion of EEG and EM, and compared the classification performance differences of the two fusion methods, the effective and powerful fusion method was Hidden Layer Fusion method.

In the study of Zhang et al. [27], explored simultaneously from a physiological and behavioral perspectives and fused pervasive electroencephalography(EEG) and audio signals to make depression more objective, effective, and convenient to detect. After extracting several effective features for these two types of signals, we trained six representative classifiers for each form, then used a co- determination tensor to correlate with the diversity of decisions of different classifiers, and combine these decisions into ultimate classification results using multiple agents. Experiments on 170 (81 depression patients and 89 normal controls) demonstrated that the proposed multi-mode depression detection system outperforms single-mode classifiers or other common late fusion strategies in accuracy, f1-scores, and sensitivity.

Shen et al. [28] proposed a multimodal-based deep learning model for automatic depression diagnosis using EATD-Corpus, a publicly available Chinese depression dataset consisting of audio and text records extracted from 162 volunteers' interviews, and DAIC-WoZ dataset, an interview-type clinical depression data. The proposed model fuses audio and text features using a Gate Recurrent Unit (GRU) model and a Bi-LSTM model with an attention layer. Text features are extracted by projecting sentences into high-dimensional sentence embeddings using ELMo. For audio features, Mel spectrogram is extracted from audio. Experimental results show that the proposed method is very effective.

Guo et al. [29] proposed a detection framework for detecting student mental health named educational data fusion for mental health detection (CASTLE). This framework is largely divided into three parts. First, using Presentation Learning, we fuse data such as social life, academic performance, and appearance. The name Multi-View Social Network Embedding (MOON), an algorithm, is proposed to effectively fuse students' disparate social relationships to express their social lives comprehensively. Second, the Synthetic Minority Oversampling Technique (SMOTE) algorithm is applied to the label imbalance problem. Finally, we use the Deep Neural Network (DNN) model for final detection. Extensive results show the promising performance of the proposed method compared to a wide range of state-of-the-art baselines.

Park et al. [30] proposed a multi-modal data-based attention-mechanism depression diagnosis model to improve the low accuracy problem of depression detection model based on single mode data. The proposed model was composed of a bidirectional encoder representations from transformers-convolutional neural network (BERT-CNN) fusion for natural language analysis, a CNN-Bidirectional Long Short-Term Memory (CNN-BiLSTM) for voice signal processing, multi-modal analysis, and a fusion model for depression diagnosis. The model used audio and text data of the Distress Analysis Interview Corpus Wizard of Oz (DAIC-WOZ) database, converted speech data into log-mel spectrograms, extracted features through the CNN model, and the extracted features were learned through the Bi-LSTM, with application of an attention mechanism. The text data was converted into embedding vectors using a BERT tokenizer, and the pre-trained BERT-CNN model was fine-tuned and trained to extract feature vectors. The proposed model resolved the problem of rapid loss increase due to the use of single mode data, and showed improved accuracy.

Xiao et al. [31] proposed a novel approach for automatic depression detection based on audio-text sequences. They proposed, as a new model for depression analysis, an attention mechanism to casual-CNN (Attention-C-CNN) for audio feature extraction. For text feature extraction, BERT, a pre-trained model proposed by Google in 2018, was used. Furthermore, a new co- attention encoder that allows improved fusion of audio and text features was applied by exchanging key-value pairs of the multi-head attention mechanism in the co- attention transformation layer. As a result of experiments using the DAIC-WOZ dataset, the proposed model showed more competitive performance than single-mode data and state-of-the-art multi-modal data-based methods.

III. PROPOSED WORK

The algorithm for sentiment prediction for stress detection is mentioned below:

Step 1: Collect the Stress data using Questionnaires and Interviews.
Step 2: Pre-process the data to remove noise and other artifacts that can affect the quality of the data.
Step 3: Plot a Wave plot to identify the relation of Amplitude and time.
Step 4: Generate a Spectrogram Visualization of the audio waveform representing magnitude spectra of consecutive frames.
Step 5: Split the Data into training and testing.
Step 6: Detect and classify the data using various ml models.
Step 7: Create an ANN model and supply data to model and predict.
Step 8: Compare DL and ML Based models.

4.3 Spectrogram Analysis

A spectrogram may be a visual representation of the spectrum of frequencies of a flag because it changes with time. When connected to a sound flag, spectrograms are some of the time called sonographs, voiceprints, or voice grams. When the information is spoken to in a 3d plot they may be called waterfalls. Spectrograms are utilized broadly within the areas of music, sonar, radar, and discourse handling, seismology, and others. Spectrograms of sound can be utilized to distinguish talked words phonetically, and to examine the different calls of creatures. A spectrogram can be produced by an optical spectrometer, a bank of band-pass channels, by Fourier change or by a wavelet change (in which case it is additionally known as a scaleogram). A spectrogram is as a rule portrayed as a warm outline, i.e., As an image with the concentrated appeared by varying the colour or brightness. Making a spectrogram utilizing the fft could be a advanced handle. Carefully inspected information, within the time space, is broken up into chunks, which as a rule cover, and Fourier changed to calculate the greatness of the recurrence range for each chunk. Each chunk at that point compares to a vertical line within the picture; a estimation of size versus recurrence for a particular minute in time (the midpoint of the chunk). These spectrums or time plots are at that point ”laid side by side” to make the picture or a three-dimensional surface,[4] or marginally covered in different ways, i.e. Windowing. This prepare basically compares to computing the squared greatness of the short time Fourier change (stft) of the flag s (t) that is, for a window width ω, spectrogram (t, ω) =|STFT (t, ω)|^2 spectrogram (t, ω) =|STFT (t, ω)|^2.

4.3.1 Sound splitting

The data-set contains 192 sound sessions between a enlivened virtual questioner called ellie, and the member. The highlights of sound portions of the members are valuable for classification, the portions are part by quiet evacuation and after that isolated by speaker diarization. These sound portions are of varying lengths such that there's spread in information.

4.3.2 Data imbalance

In the data-set, the number of non-depressed subjects is around four times bigger than that of discouraged ones, which can present a classification ”non-depressed” inclination. Extra inclination can happen due to the impressive run of meet lengths from 7-33 minutes since a bigger volume of flag from a person may emphasize a few characteristics that are individual specific. To correct this lopsidedness, sound fragments are haphazardly tested in break even with numbers.

4.3.3 Spectrogram Conversion

The examined sound portions are at that point changed over to spectrogram pictures of estimate 512 x 512 pixels. These pictures are put into distinctive envelopes comparing diverse classes and after that the envelopes are part into preparing and approval information with proportion 8:2.

V. RESULTS

Table 6.1 below show the comparison of the existing and proposed models. The results are evaluated in terms of accuracy.

Implemented Algorithm	Results in Accuracy
Decision Tree	56%
Random forest	61%
Proposed CNN	67%

Table 6.1: Accuracy Comparison.

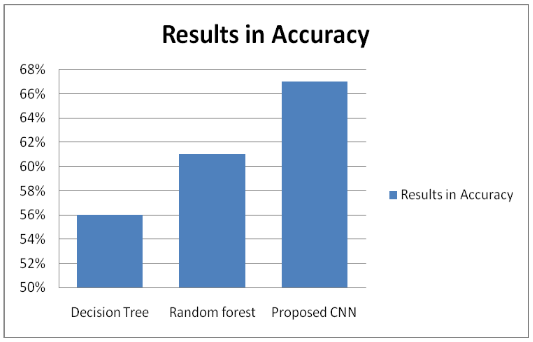

Figure 6.10: Accuracy Chart Comparison.

VI. CONCLUSION

In this paper, we designed a audio-stream-based depression diagnosis model using the late score fusion method for CNN models based on the use of audio data and compared the performance of the proposed model. Depression diagnosis using multi-modal data allows acquiring more information for depression, a complex mental health condition, and thus is effective for diagnosis. Among audio data, linguistic features such as speech reveal the characteristics of people with symptoms of depressive disorder.

REFERENCES

[1] L. R. Nair, S. D. Shetty, and S. D. Shetty, ‘‘Applying spark based machine learning model on streaming big data for health status prediction,’’ Comput. Electr. Eng., vol. 65, pp. 393–399, Jan. 2018, doi: 10.1016/j.compeleceng.2017.03.009.

[2] Apache Spark. (2018). Apache Spark TM—Unified Analytics Engine for Big Data. [Online]. Available: https://spark.apache.org.

[3] Apache Hadoop. Accessed: Oct. 26, 2022. [Online]. Available: https://hadoop.apache.org/.

[4] Apache Storm. Accessed: Oct. 26, 2022. [Online]. Available: https://storm.apache.org/.

[5] Apache Kafka. Accessed: Oct. 27, 2022. [Online]. Available: https://kafka.apache.org/.

[6] A. M. Y. Tai, A. Albuquerque, N. E. Carmona, M. Subramanieapillai, D. S. Cha, M. Sheko, Y. Lee, R. Mansur, and R. S. McIntyre, ‘‘Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry,’’ Artif. Intell. Med., vol. 99, Aug. 2019, Art. no. 101704, doi: 10.1016/j.artmed.2019.101704.

[7] D. Bzdok and A. Meyer-Lindenberg, ‘‘Machine learning for precision psychiatry: Opportunities and challenges,’’ Biol. Psychiatry, Cogn. Neurosci. NeuroImage. vol. 3, no. 3, pp. 223–230, Mar. 2018, doi: 10.1016/j.bpsc.2017.11.007.

[8] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh, “An inventory for measuring depression,” Arch. Gen. Psychiatry, vol. 4, no. 6, pp. 561–571, 1961.

[9] S. H. Hosseini-Saravani, S. Besharati, H. Calvo, and A. Gelbukh, “Depression detection in social media uses a psychoanalytical technique for feature extraction and a cognitive based classifier,” in Proc. Mex. Int. Conf. Artif. Intell. Springer, 2020, pp. 282–292. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3- 030-60887-3_25.

[10] S. Rude, E.-M. Gortner, and J. Pennebaker, “Language use of depressed and depression-vulnerable college students,” Cogn. Emotion, vol. 18, no. 8, pp. 1121–1133, Dec. 2004.

[11] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, “Detection of depression related posts in reddit social media forum,” IEEE Access, vol. 7, pp. 44883–44893, 2019.

[12] S. Ji, X. Li, Z. Huang, and E. Cambria, “Suicidal ideation and mental disorder detection with attentive relation networks,” Neural Comput. Appl., pp. 1–11, Jun. 2021. [Online]. Available: https://link.springer. com/article/10.1007/s00521-021-06208-y.

[13] M. Paul and M. Dredze, “You are what you tweet: Analyzing Twitter for public health,” in Proc. Int. AAAI Conf. Web Social Media, 2011, vol. 5, no. 1, pp. 265–272.

[14] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, “Characterizing and predicting postpartum depression from shared Facebook data,” in Proc. 17th ACM Conf. Comput. Supported Cooperate. Work Social Comput. Feb. 2014, pp. 626–638.

[15] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with Twitter data,” Sci. Rep., vol. 7, no. 1, pp. 1–11, Dec. 2017.

[16] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, “SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis,” in Proc. 29th ACM Int. Conf. Inf. Knowl. Manage. Oct. 2020, pp. 105–114.

[17] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms. Hoboken, NJ, USA: Wiley, 2014.

[18] Depression and Other Common Mental Disorders: Global Health Estimates, World Health Org., Geneva, Switzerland, Tech. Rep., 2017.

[19] Y. Huang et al., “Prevalence of mental disorders in China: A cross-sectional epidemiological study,” Lancet Psychiatry, vol. 6, no. 3, pp. 211–224, 2019.

[20] C. F. Reynolds et al., “Early intervention to reduce the global health and economic burden of major depression in older adults,” Annu. Rev. Public Health, vol. 33, no. 1, pp. 123–135, Apr. 2012.

[21] F. Edition et al., “Diagnostic and statistical manual of mental disorders,” American Psychiatric Association, vol. 21, no. 21, pp. 591–643, 2013.
[22] M. Hamilton, “The Hamilton rating scale for depression,” in Assessment of Depression. Cham, Switzerland: Springer, 1986, pp. 143–152.
[23] A. T. Beck, R. A. Steer, and G. Brown, “Beck depression inventory—II,” Psychol. Assessment, Jan. 1996.
[24] K. Kroenke and R. L. Spitzer, “The PHQ-9: A new depression diagnostic and severity measure,” Psychiatric Ann., vol. 32, no. 5, pp. 509–515, 2002.
[25] B. Ay, O. Yildirim, M. Talo, U. B. Baloglu, G. Aydin, S. D. Puthankattil, and U. R. Acharya, ‘‘Automated depression detection using deep representation and sequence learning with EEG signals,’’ J. Med. Syst., vol. 43, no. 7, pp. 1–6, Feb. 2019.
[26] J. Zhu, Y. Wang, R. La, J. Zhan, J. Niu, S. Zeng, and X. Hu, ‘‘Multimodal mild depression recognition based on EEG-EM synchronization acquisition network,’’ IEEE Access, vol. 7, pp. 28196–28210, Feb. 2019.
[27] X. Zhang, J. Shen, Z. U. Din, J. Liu, G. Wang, and B. Hu, ‘‘Multimodal depression detection: Fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble,’’ IEEE J. Biomed. Health Informat., vol. 23, no. 6, pp. 2265–2275, Nov. 2019.
[28] Y. Shen, H. Yang, and L. Lin, ‘‘Automatic depression detection: An emotional audio-textual corpus and a GRU/BiLSTM-based model,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2022, pp. 6247–6251.
[29] T. Guo, W. Zhao, M. Alrashoud, A. Tolba, S. Firmin, and F. Xia, ‘‘Multimodal educational data fusion for students’ mental health detection,’’ IEEE Access, vol. 10, pp. 70370–70382, 2022.
[30] J. H, Park and N. M. Moon, ‘‘Design and implementation of attention depression detection model based on multimodal analysis,’’ Sustainability, vol. 14, no. 6, pp. 1–15, 2022.
[31] J. Xiao, Y. Huang, G. Zhang, and W. Liu, ‘‘A deep learning method on audio and text sequences for automatic depression detection,’’ in Proc. 3rd Int. Conf. Appl. Mach. Learn. (ICAML), Jul. 2021, pp. 388–392.