# *Multimodal Depression Detection System Using Machine Learning*


Govind

*Department of Information Technology*
*JSS Academy of Technical Education*
Noida, India
saraswatgovind27@gmail.com

Pranav Arya

*Department of Information Technology*
*JSS Academy of Technical Education*
Noida, India
pranav.arya158@gmail.com

Gunjan Ansari

*Department of Information Technology*
*JSS Academy of Technical Education*
Noida, India
gunjanansari@jssaten.ac.in

Yash Saxena

*Department of Information Technology*
*JSS Academy of Technical Education*
Noida, India
yashsaxena427@gmail.com

Alok Sharma

*Department of Information Technology*
*JSS Academy of Technical Education*
Noida, India
19it050@jssaten.ac.in



2023 Second International Conference on Informatics (ICI) | 979-8-3503-4383-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICI60088.2023.10421362



***Abstract*— Mental health is a crucial aspect of one's overall well-being, and identification of depression at an early stage is crucial in prevention of complications at a later stage. In recent years, depression is a major health concern which has led many researchers to devise automated tools and techniques that can assist medical practitioners in depression detection. Previous studies on depression detection focused mainly on only textual mode of data leading to less accurate results. The proposed work use natural language processing and computer vision techniques to analyze the audio, video, and transcript data to discriminate between depressed and non-depressed people. The features such as MFCCs and entropy are extracted from audio data, Facial expressions, Augmentation Units (AUs), and gaze features are extracted from videos and lexical and syntactic features are extracted from transcript data. The extracted features from all three modes are combined using feature-based fusion and Principal Component Analysis is applied further to select relevant features to improve model accuracy. The experimental study is conducted on DAIC-WOZ dataset that consists of speech, facial expression, and physiological signals of individuals diagnosed with depression and a control group without depression. The widely used machine and deep learning models such as Logistic Regression (LR), Naïve Bayes, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) are employed to classify individuals as depressed or non-depressed. The results indicate that the feature-level fusion of all three modes can significantly improve the performance of the models as compared to unimodal and bi-modal fusion. The best f1-score of around 98% is achieved when PCA is utilized for feature selection and Logistic Regression is employed as classifier.**



**Keywords— Multimodal data, Depression Detection, DAIC-WOZ, Principal Component Analysis**


## I. INTRODUCTION

According to the World Health Organization, approximately 264 million people suffer from depression globally, making it a significant public health concern. Identifying and treating depression early on is essential in preventing further complications and improving the quality of life for those affected.



Machine learning techniques have shown promising results in identifying and predicting depression. The authors in [1] employed machine learning algorithms to predict severity levels of depression, anxiety, and stress (DAS) using the score achieved in the set of questionnaires based on the Depression, Anxiety, and Stress Scale (DASS). Another domain that has gained attention in recent years is the detection of Depression and Anxiety though social media (DASM) posts collected from various social media platforms. The research in [2] employed machine learning algorithms such as Naïve Bayes and a hybrid model, NBTree to classify the user into depressed and non-depressed groups based on the collected data from Twitter.

Multimodal depression detection has also gained popularity in the recent years due to better accuracy than unimodal data. A Chinese Multimodal structure is created in [3] which consists of Audio, visual and transcript data of individuals suffering from Clinical Depression. One more popular dataset for depression detection was DAIC-WOZ (The Distress Analysis Interview Corpus Wizard-of-Oz) that consists of clinical interviews of individuals categorized into depressed and non-depressed. The dataset was created for the diagnosis of psychological disorders such as anxiety, depression, and stress disorder. The dataset contains speech, facial expression, and physiological signals of individuals diagnosed with depression and a control group without depression. In one of the recent works in [4], authors proposed a two-layered model for multi-modal depression detection on DAIC-WOZ dataset that extracts features from individual responses at interview questions and then identify their semantic categories.

The significance of this proposed work lies in the potential to provide a cost-effective and accessible method for early detection and intervention of depression. Early detection of depression can significantly reduce the economic and social burden associated with the disorder. Furthermore, the use of machine learning models in mental health research has the potential to provide valuable insights into the underlying mechanisms of depression and improve current treatment approaches.

The key contribution of this proposed work is to develop a machine learning model for improving the accuracy of

depression detection using feature-level fusion of multimodal data. In the proposed work, PCA is utilized to select relevant and non-redundant features at the fusion layer. The widely used machine and deep learning models such as Logistic Regression (LR), Naïve bayes, Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) are employed to classify individuals as depressed or non-depressed. The models' performance is evaluated on the DAIC-WOZ dataset using performance metrics, including accuracy, precision, recall, and F1-score. The results indicate that the combination of features from all three modes can significantly improve the performance of the model. The best f1-score of around 98% is achieved when PCA is utilized for feature selection and Logistic Regression is employed as classifier.

## II. Literature Review

Several studies have been conducted to explore the application of multimodal analysis for depression detection. This section discusses relevant research papers that have investigated different approaches and techniques for detecting depression using multimodal data. In the study by Rosas, Mihalcea, and Morency [5], the focus was on multimodal sentiment analysis of online Spanish videos. Although not specifically targeting depression detection, this research demonstrated the potential of analyzing multiple modalities, including audio, video, and text, to capture emotional states. The authors utilized transcriber software, General Inquirer software, and tools like OpenEAR and Praat for audio and textual analysis.

Yalamanchili [6] proposed a real-time acoustic-based depression detection system using machine learning techniques. They explored the use of acoustic features extracted from audio recordings to detect depression accurately. This study highlighted the significance of incorporating audio modalities in depression detection. Yoon, Kang, Kim, and Han [1] conducted research on detecting depression through the analysis of vlogs. They created the D-vlog dataset, which facilitated multimodal analysis of audio, visual, and textual features. By integrating these modalities, they aimed to enhance understanding and detection of depression indicators. The study reported improved accuracy by combining textual and visual data.

These studies collectively highlight the importance of multimodal analysis in depression detection. By integrating audio, visual, and textual modalities, researchers have demonstrated improved accuracy in capturing depression indicators and understanding the emotional states of individuals. The methodologies employed in these studies, such as deep learning architectures, feature extraction techniques, and fusion strategies, provide valuable insights for developing robust and effective depression detection systems.

## III. Proposed Methodology

Fig. 3.1 shows the complete architecture of the proposed work, and the following subsections discuss the details of the various modules employed in this work at different phases.

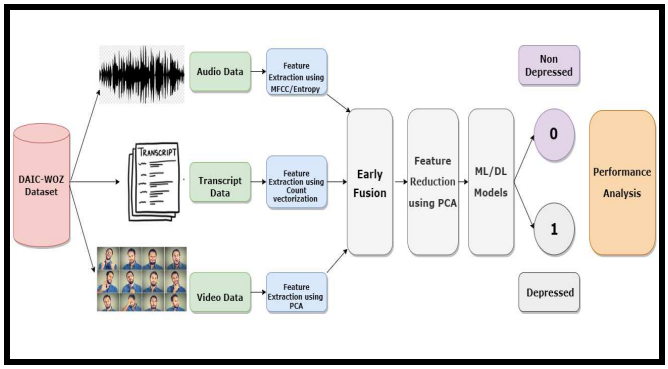


**Fig 3.1: Proposed architecture of Multimodal Depression Detection**

### A. Dataset Used

The dataset used in the proposed work is the Distress Analysis Interview Corpus/Wizard-of-Oz (DAIC- WOZ) [7], [8]. The dataset consists of clinical interviews of 189 subjects, out of which only 169 subjects are used for the analysis. The dataset is skewed as it has 124 depressed and 40 controlled subjects. The dataset comprises of approximately 189 folders of sessions numbered 300-492. Certain sessions have been excluded for technical reasons. Each candidate's information is stored in a dedicated folder, containing files such as CLNF_features.txt, CLNF_hog.bin, CNLF_gaze.txt, CNLF_AUs.csv, AUDIO.wav, COVAREP.csv, CLNF_pose.txt, FORMANT.csv, and TRANSCRIPT.csv. The AUDIO.wav file consists of the audio recording of the conversation, while COVAREP.csv contains information about the audio features of the patient. TRANSCRIPT.csv contains textual data analysis, while the remaining files represent video features of the individual.

### B. PRE-PROCESSING STEPS

As the multimodal consists of audio, video and transcript data, different pre-processing methods are employed according to the type of data. In the pre-processing of textual data, cleaning of the data is performed to remove noise, followed by the tokenization of the data to create a list of words present in the sentences spoken by the speaker. The use of stemming and stop words is not feasible as these feature also play a role in the domain of study and removing them will negatively affect the prediction. The technique of count vectorization is employed where for a given statement X, its count vector X (i, j) indicates the count of word i occurring j times in the statement X [1], [4], [5].

For audio pre-processing, audio spectrum, entropy of the signal waveform is calculated and used with other audio features such as Mel Frequency Cepstral Coefficients (MFCCs) [2], [4]. Spectrogram is an audio application that

comes through Fourier Series which gives a sound wave representation as shown below [5].

$$X[k] = \sum_{m=0}^{M-1} \left(x[m] * e^{\frac{-j2\Pi k*m}{M}}\right) \quad (1)$$

where, x[m] represents the input sequence of length M, X[k] represents the complex-valued output spectrum at frequency bin k, j is the imaginary unit (√(-1)), k is the index of the frequency bin ranging from 0 to M-1.

$$C(n) = \sum_{k=0}^{M-1} \left(\log E(k) * \cos\left(\Pi * n * \frac{k+0.5}{M}\right)\right) \quad (2)$$

where, C(n) represents the n-th MFCC coefficient, log E(k) represents the logarithm of the energy of the k-th Mel filter output, M is the total number of Mel filter banks.

Entropy on the other hand quantifies the complexity of an audio signal and is computed using the Shannon entropy formula as shown below.

$$H = -\sum_{i=0}^{N} \left(P[i] * log_2(P(i))\right) \quad (3)$$

Where, H represents the entropy of the signal, N is the number of possible values or bins in the signal, P(i) represents the probability of occurrence of the i-th value or bin.

The features of subjects extracted from video files consist of Action Units (AUs), gaze, Histogram of Oriented Gradients (HoG), and Constrained Local Neural Fields (CLNF). AUs are extracted using a facial action unit detection algorithm that analyses the shape of the face in each frame of the video to identify the presence and intensity of each AU. Gaze is extracted using an eye tracking algorithm that is used to track the movement of the eyes in each frame of the video to identify the direction of gaze. HOG features are extracted using a histogram of oriented gradients algorithm. CLNF features are extracted with the help of convolutional neural network. This model is trained on a large volume of textual data to learn how to represent words and phrases as in [4]. After feature extraction from all modes, data is transformed on a similar scale using Min-Max normalization. This is performed to improve the stability of the models while training, thus improving its overall performance.

### C. FUSION

In the proposed work, early fusion is applied, which involves combining the various features belonging to single mode of data. The fusion technique is used to obtain an improved feature set for model training that utilizes the advantages of features extracted from all three different modes of data. Bi-modal and tri-modal fusion are performed, and the results are analyzed based on the early fusion approach [4].

### D. PRINCIPAL COMPONENT ANALYSIS

Fusion leads to better accuracy but it brings in a very large pool of features for training. To overcome this issue, PCA is used as a feature reduction technique so that relevant information can be gathered to improve model accuracy. PCA works by identifying the directions, called principal components as explained in [2], along which the data varies the most. These principal components are orthogonal to each other and are ranked in order of their significance. The first principal component captures the highest value of variance in the data, followed by the second component, and so on.

### E. MACHINE LEARNING AND DEEP LEARNING CLASSFIERS

In this work, four popular machine/deep learning models are employed. The first classifier used is Logistic regression, which is a statistical model that is used for binary classification of data. The logistic regression model assumes a logistic or sigmoidal relationship between the input variables and the probability of the positive class. This in turn evaluates a value that can either be 0 or 1, acknowledging whether this data belongs to that specific class or not [2], [3], [4]. Another algorithm used is Naïve Bayes classifier based on the Bayes' theorem, which computes the conditional probability of an event based on prior knowledge [3]. Another popular classifier SVM is employed in this work as previous studies show that it achieves better generalization and robustness in classifying unseen data. The algorithm searches for the optimal hyperplane that separates the data points of different classes with the largest margin [2]. This hyperplane is chosen to maximize the distance between the nearest data points of each class, known as support vectors [3]. The last classifier employed for depression detection in this work is a deep learning model- Long Short-Term Memory (LSTM) as in [2], [3] that represents a type of recurrent neural network (RNN) architecture. It is specifically designed to address the limitation of the vanishing gradient problem that occur in RNN when they attempt to capture such long-term dependencies in sequential data characterized by patterns over time, such as text, speech, and time series data.

## IV. EXPERIEMENTAL RESULTS

In this section, experimental settings used for implementation of the proposed model are discussed. The section also shows the result of different learning algorithms with unimodal, bimodal and tri modal data with and without feature selection techniques.

### A. EXPERIEMNTAL SETUP

In the proposed work, different combinations of modal fusions and models with varying hyper parameter settings are explored to optimize the performance of depression detection. Hyperparameters are optimized through techniques such as Grid Search CV, use of linear and sigmoid kernels in SVM and varying the dropout rate, number of hidden neurons and optimizers in LSTM. LSTM achieves best accuracy with dropout rate=0.3, number of hidden neurons=32 and Adam Optimizer. Scikit learn is employed for implementation of machine learning models, Tensorflow and keras for deep learning and Librosa is used for audio pre-processing.

### B. DATA PRE-PROCESSING AND TRANSFORMATION

In this proposed work, three different modes of data are used, they are transcript, audio and video. The data from all these modes is pre-processed according to their type and transformed into a vector for the machine learning models. In transcript data, the count vector of size 3000 is acquired which represents the frequency of 3000 most spoken words by a participant. In audio data, two features - entropy and MFCC corresponding to the audio of a participant are extracted. The pre-processing of video data gives various visual features regarding gaze, pose and facial action units which is represented as a vector.

### C. RESULTS

The experimental study is conducted on various combinations of data with and without feature selection on the dataset and performance is evaluated using accuracy and f1-score. Accuracy is computed to predict models' overall performance whereas f1 measure is computed to prove that the system is more efficient in terms of precision and recall. F1 score is computed as there is non- uniform distribution of depressed and non-depressed samples. The results shows accuracy in percentage and f1 score in the range of 0-1. To show the effect of normalization on the dataset, the performance is evaluated on both normalized and non-normalized data.

Table I demonstrates the impact of modalities and classifiers on non-normalized data in terms of accuracy. It is observed that bimodal and trimodal data consistently outperform unimodal data. The combination of video and transcript data enhances accuracy to 90.56%. Trimodal data yields results similar to bimodal data in the non-normalized form.

Table II describes the f1-score of different modalities for non-normalized data. In this it's clearly observed that the f1-score of bi-modal and tri-modal is higher than the uni-modal.

Table III shows the result of all modalities and classifiers on normalized data in terms of accuracy. As observed from the table, bimodal and trimodal data consistently outperform unimodal data. Combining audio and transcript data achieves 96% accuracy (SVM-Linear). Trimodal data exhibits the best performance in the normalized form, reaching 98% accuracy.

Table IV describes the f1-score of different modalities for normalized data. In this it's clearly observed that the f1-score of bi-modal and tri-modal is higher than the unimodal.

Table V also demonstrates that bimodal and trimodal data outperform unimodal data in terms of accuracy. Combining video and transcript data increases accuracy to 90% (SVM-sigmoid), while transcript and audio fusion achieves 96% accuracy (Logistic Regression) with an improved. Trimodal data shows the best performance with 98% accuracy.

Table VI shows the f1- score of different modalities. It can be observed that by using bi-modal and tri-modal the f1-score is better than uni-modal.

### D. COMPARISON WITH STATE-OF-THE-ART

Unimodal or bimodal approaches were utilized in previous studies on depression detection, yielding accuracies of up to 93%. However, these approaches are subject to limitations. Unimodal approaches solely focus on one data mode, while bimodal approaches are constrained by the independence of data modes. The multi-modal model proposed in this study integrates visual, audio, and textual data, resulting in an improved accuracy of 90%. This comprehensive approach enhances accuracy and holds promise for clinical applications in depression diagnosis and early intervention. Table VII shows comparison between previous works and the proposed work in terms of performance metrics on the DAIC-WOZ employed in the proposed work.

As observed from Table VII, machine learning models such as SVM perform better than LSTM in almost all studies. The reason for this poor performance of deep learning models can be availability of a smaller number of data training samples and data skewness. Although the work by Bhanusree et al. achieved remarkable performance of 93% after balancing data using SMOTE, our proposed work outperforms it with 98% accuracy score. It can be concluded that the use of data normalization, feature selection and Grid CV in SVM facilitated in achieving better performance as compared to other works.

## VI. Conclusion

The aim of this proposed work is to improve upon the accuracy achieved by previous studies in the domain of depression detection. This objective was accomplished by implementing a multi-modal approach that surpasses the performance of both unimodal and bimodal approaches. The proposed method combines visual, audio, and textual data, followed by feature selection and extraction of relevant attributes, resulting in better performance in terms of both accuracy and f1 score. The use of the DAIC-WOZ dataset, which encompasses multiple data modes, facilitates a comprehensive analysis and enhances performance across different scenarios. The result analysis depicts that through the use of hyperparameter tuning and data normalization, there is significant improvement in model accuracy. The superiority of bimodal and trimodal data over unimodal data is consistently observed, highlighting the benefits of leveraging multiple modes of information for depression detection. Remarkably, the highest accuracy is attained with

## References


[1] J. Yoon, C. Kang, S. Kim, and J. Han, "D-vlog: Multimodal vlog dataset for depression detection," in *AAAI Conference on Artificial Intelligence*, June 2022, vol. 36, no. 11, pp. 12226-12234.

[2] C. Zhang, "Based on Multi-Feature Information Attention Fusion for Multi-Modal Remote Sensing Image Semantic Segmentation," in 2021 *IEEE International Conference on Mechatronics and Automation (ICMA),* Takamatsu, Japan, 2021, pp. 71-76.

[3] B. Zou et al., "Semi-structural Interview-Based Chinese Multimodal Depression Corpus Towards Automatic Preliminary Screening of Depressive Disorders," *IEEE Transactions on Affective Computing,* 2022, pp. 1-16, doi: 10.1109/TAFFC.2022.3181210.

[4] S. Guohou, Z. Lina, and Z. Dongsong, "What reveals about depression level? The role of multimodal features at the level of interview questions," *Information & Management,* vol. 57, no. 7, pp. 103349, 2020.

[5] V. P. Rosas, R. Mihalcea, and L. P. Morency, "Multimodal sentiment analysis of Spanish online videos," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 38-45, 2013.

[6] B. Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella and S. V. Alluri, "Real-time Acoustic based Depression Detection using Machine Learning Techniques," in 2020 *International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, India, 2020, pp. 1-6.

[7] A. Elgammal and R. Kiros, "Depression detection from dialogue using multimodal deep learning," in *Empirical Methods in Natural Language Processing and Applications (EMNLP- IJCNLP),* 2018, pp. 7780-7789.

[8] M. Garg, "Mental health analysis in social media posts: a survey," *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1819-1842, 2023.

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[10] C. H. Cai, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, and B. Hu, "MODMA dataset: a multi-modal open dataset for mental-disorder analysis," *arXivpreprint arXiv:2002.09283*, 2020.

[11] 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'14),* Paris, 2014 pp. 1061-1068.

[12] R. P. Thati, A. S. Dhadwal, and P. Kumar, "A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms," *Multimedia Tools and Applications*, vol. 82, no. 4, pp.1-34, 2022.

[13] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, Jr., "A Very Brief Measure of the Big Five Personality Domains," *Journal of Research in Personality*, vol. 37, pp. 504-528, 2003.

[14] J. Gratch et al., "The Distress Analysis Interview Corpus of human and computer interviews," in *LREC*, 2014, pp. 3123-3128.

[15] K. A. Govindasamy and N. Palanichamy, "Depression Detection Using Machine Learning Techniques on Twitter Data," in *5th International Conference on Intelligent Computing and Control Systems (ICICCS),* Madurai, India, 2021, pp. 960-966.

[16] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PloS One*, vol. 13, no. 5, pp. e0196391, 2018.


Table I. Accuracy of different classifiers (non-normalized data)

| Modalities | SVM-sigmoid | Naïve Bayes | LR | LSTM |
|---|---|---|---|---|
| Audio | 70.17 | 68.42 | 70.17 | 68.31 |
| Video | 74.51 | 59.32 | 68.62 | 68.42 |
| Text | 71.05 | 71.79 | 74.78 | 73 |

| | | | | |
|---|---|---|---|---|
| Video + Text | 74 | 90.56 | 72 | 71.43 |
| Audio + Text | 90 | 90 | 88 | 72.4 |
| Audio + Video + Text | 90 | 90.56 | 86 | 77.47 |

Table II. F1-score of different classifiers (non-normalized data)

| Modalities | SVM-sigmoid | Naïve Bayes | LR | LSTM |
|---|---|---|---|---|
| Audio | 0.57 | 0.52 | 0.57 | 0.56 |
| Video | 0.58 | 0.63 | 0.63 | 0.59 |
| Text | 0.76 | 0.69 | 0.67 | 0.61 |
| Audio + Video | 0.54 | 0.52 | 0.63 | 0.71 |
| Video + Text | 0.62 | 0.88 | 0.65 | 0.65 |
| Audio + Text | 0.86 | 0.88 | **0.89** | 0.62 |
| Audio + Video + Text | 0.86 | 0.88 | 0.87 | 0.69 |

Table III. Accuracy of different classifiers (Normalized data)

| Modalities | SVM-sigmoid | Naïve Bayes | LR | LSTM |
|---|---|---|---|---|
| Audio | 70.17 | 68.42 | 70.17 | 68.31 |
| Video | 74.51 | 59.32 | 68.62 | 68.42 |
| Audio + Video | 68.29 | 63.07 | 62.31 | 74.51 |
| Video + Text | 81 | **90.56** | 86.27 | 72.34 |
| Audio + Video + Text | 90 | **90.56** | 86.79 | 77.15 |

Table IV. F1-score of different classifiers (Normalized data)

| Modalities | SVM-sigmoid | Naïve Bayes | LR | LSTM |
|---|---|---|---|---|
| Audio | 0.57 | 0.67 | 0.57 | 0.61 |
| Video | 0.58 | 0.62 | 0.63 | 0.58 |
| Audio + Video | 0.55 | 0.57 | 0.64 | 0.59 |
| Video + Text | 0.72 | 0.81 | 0.89 | 0.67 |
| Audio + Text | 0.86 | 0.82 | 0.81 | 0.69 |
| Audio + Video + Text | 0.86 | 0.82 | **0.98** | 0.71 |

Table V. Accuracy of different classifiers with PCA

| Modalities | SVM-sigmoid | Naïve Bayes | LR | LSTM |
|---|---|---|---|---|
| Video | 0.6 | 62.76 | 67.5 | 0.68 |
| Audio + Video | 51.28 | 69.23 | 71.79 | 74.5 |
| Video + Text | 90 | 67.92 | 78 | 73.43 |
| Audio + Text | 92 | 75.47 | 96 | 72.98 |
| Audio + Video + Text | 90 | 71.69 | **98** | 75.57 |

Table VI. F1-score of different classifiers with PCA

| Modalities | SVM-sigmoid | Naïve Bayes | LR | LSTM |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Video | 0.57 | 0.57 | 0.65 | 0.68 |
| Audio + Video | 0.59 | 0.63 | 0.65 | 0.71 |
| Video + Text | 0.85 | 0.72 | 0.86 | 0.69 |
| Audio + Text | 0.90 | 0.72 | 0.96 | 0.71 |
| Audio + Video + Text | 0.86 | 0.76 | **0.98** | 0.72 |

Table VII. Comparison with the existing works

| **Authors** | **Methodology used** | **Best Accuracy achieved** |
|---|---|---|
| Ashish Nayak et al.[7] | CNN, ANN, LSTM and SMOTE for data imbalance | LSTM : 76.27% |
| Muhammad Muzammela et al.[4] | CNN & LSTM with fusion of audio+ video+textual features | LSTM on Trimodal data: 78.4% |
| Bhanusree Yalamanchili et al.[4] | LR, RF and SVM with COVAREP, SMOTE & PCA | SVM after SMOTE: 93% |
| Ravi Prasad Thati et al.[5] | LR, Naive Bayes, SVM, decision Tree, Random Forest | SVM : 86 % |
| Proposed Work | SVM, Naïve Bayes, Logistic Regression & LSTM with PCA for feature reduction | SVM(linear) and LR: 98% |

Table VIII. LSTM Models used.

| Model Type | Number of Layers | Units in Layer | Dense Layer units | Dense Layer Activation function |
|---|---|---|---|---|
| Sequential | 1 | 128 | 1 | relu |
| Sequential | 1 | 64 | 1 | sigmoid |
| Sequential | 1 | 128 | 1 | linear |