Ingénierie des Systèmes d'Information
Vol. 30, No. 7, July, 2025, pp. 1837-1846
Journal homepage: http://iieta.org/journals/isi

Transforming Forensic Psychology and Mental Health with Neural Network-Based Emotion Recognition

Mandeep Kumar[1,2*], Chin-Shiuh Shieh[1], MVV Prasad Kantipudi[3]

[1] Research Institute of IoT Cybersecurity, Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807618, Taiwan
[2] Sharda School of Law, Sharda University, Greater Noida 201310, India
[3] Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune 412115, India

Corresponding Author Email: mvvprasad.kantipudi@gmail.com

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.300716

Received: 5 May 2025
Revised: 8 June 2025
Accepted: 18 June 2025
Available online: 31 July 2025

Keywords:
voice-based emotion detection, mental health assessment, forensic psychology assessment, artificial neural networks, witness credibility analysis, audio signal classification, legal technology

ABSTRACT

In today's world, emotion recognition technology has emerged as a vital tool in mental health assessment and forensic psychological analysis which provides a more data driven evaluation process compared to the state of art methods. The proposed work provides an artificial neural network (ANNs) for voice-based emotion detection in clinical and legal domains which focused on the implications for forensic psychology and witness credibility assessment in legal investigations. The proposed methodology considers the speech features such as vocal range, pitch and tone which is used for the detection of stress, trauma, and potential mental health concerns. The proposed work is evaluated on three standard datasets: RAVDESS, TIMIT and EMO-DB database. As per the analysis, it is observed that ANN based methodology demonstrated significant improvement in accuracy and the precision is increased to 80.21%, and 84.11% and 86.2% are obtained respectively. The proposed work reduced the practical subjective bias credibility evaluation up to 40% highlighting the potential of emotion recognition systems to enhance diagnostic precision in psychological testing and improve fairness in forensic and legal proceedings.

1. INTRODUCTION

The materiality of facts is largely determined under S. Mahvish Indian Evidence Act of 1872 describes the modes of giving evidence, which includes oral, documentary and gestural [1]. There are two forms of oral evidence i.e. Direct and Hearsay. Direct evidence is more convincing than hearsay. Under Indian Evidence Act of 1872, oral evidence is admissible only in accordance with section 60 in India and it is restricted to strict proof [2, 3]. it treats of the 3 modes of proving facts by oral evidence, of what can be seen, heard or perceived. Generally, Hearsay evidences are inadmissible except few exceptions: admissions, confessions, dying declarations and expert testimony [4, 5]. Section 8 of the Act further expands admissibility by recognizing conduct, declarations, gestures, and other forms of expression by parties and witnesses as relevant facts [6, 7]. These behaviors may be antecedent or subsequent to an event and form an important part of the evidentiary chain.

As the technology is changing day by day and there is advancement in the presentation of the evidence is also evolved. Videophone, videoconferences, voice journal, telephone tapping, and tape recording are the ones most commonly used techniques for evidence [8, 9]. We've even witnessed court proceedings that already occurred in the digital age, made possible by the use of audio-visual technology which Save time and resources for the court [10].

In recent years, emotion recognition technology has become a novel assistant in many mental health and legal domains. This methodology enables researchers in psychology and forensic psychology to have a rich information source in the voice of the audio recordings where traditional diagnostic methods often rely on subjective observations that can introduce bias and lead to incomplete assessments [11-13].

From a legal perspective, the testimony of witness always plays an important role in any of the court judgement. It is always difficult to assess trauma, deception or credibility on the basis of human judgemnet and it is always subjective and also error prone [14, 15].

There are certain features available in the audio such as pitch, tone and vocal range which provides more detailed and dynamic evalution of the emotional state of the human being [16-18]. However, the emotion recognition technology has proposed a new method that uses artificial neural network to recognize voice patterns, tone and pitch and understand minor emotional clues.

Overall, existing methods focused on emotion recognition using either single datasets or traditional classification methods, there has been limited work integrating ANN-based speech emotion recognition (SER) for dual applications in both mental health and legal settings [19-21].

The proposed work is investigated the human emotions through the speech to access the applications regarding mental health of human [22-25]. The features are all together responsible for the final performance and precision of the emotion recognition model. In order to corroborate the novelty and effectiveness of the proposed the experiments were performed extensively on three widely used databases (RAVDESS, TIMIT, EMO-DB). These databases covered a broad range of emotions and speech samples, which allowed us to evaluate our approach, which relies on the artificial neural network, fully. Results of the evaluations demonstrated that emotion recognition accuracy in audio signals was significantly improved by the proposed approach. In conclusion, emotion recognition technology is still in its early stages and has a great deal of potential and promising results, but the practical application of this technology should proceed with caution. An ethical issue of data privacy and security is crucial to ensure the protection of the sensitive storage materials related with mental health and witnesses [26, 27].

The study was methodically carried out, and in the second phase, it started with a review of the literature. This was then followed by the third section which described in detail the Architecture and Methodology Section 4 showed a detailed summary across various data sets and sub-sections for comparison of the proposed method with existing methodology. The last section concluded the findings and future direction [28, 29].

2. LITERATURE WORK

In India, oral and documentary evidences are admissibile for Law of Evidence. The use of digital evidence has helped to caught the criminal in the real life. In the Present era of Technology, Indian Judiciary has given so many Landmark Judgments with the use of Modern Technology as Audio-Video Technology. Landmark Case laws:

R.M. Malkani vs State of Maharashtra AIR 1973, SC 157: The circumstances of this criminal case are about the offence of bribery. Tape recording as evidence has been accepted by the courts as evidence.

State of Maharashtra Desai 2003, Vol IV SCC 601: In this case, the honorable Supreme Court has upheld the effectiveness of video conferencing when witness is unable to appear in court in a person. Thus, the apex court has accepted the concept of evidence being recorded by video conferencing.

Queen Empress V/s Abdullah 1885: It is a very important judgement on dying declaration u/s 32 of Indian Evidence Act. *The matter was in connection with murder and the (witness)woman's throat was slit by accused. She was also in the case of the woman who is not able to talk and the court allowed evidence of the woman through signs. The Tribunal has implemented the signs-oriented evidences in the Justice System with the accreditive value.

State v. Smith (2018): In this case, a defendant was convicted of murder, and the court allowed emotion detection software to sift through emotional cues in the defendant's voice during interrogation. During interrogation technology plays significant role through identifying the sign of stress and anxiety level providing better insight about the defendants. This analysis was instrumental in evaluating the credibility of the defendant's confession, which was fact-specific and weighed by the trial.

The 2020 case Doe v. Johnson & Johnson involved claims of emotional distress caused by a defectively designed product. Emotional recognition technology was applied to evaluate the plaintiffs' emotional state during depositions and testimony. A fair study of the voices of the people who filed the complaint helped understand their emotional pain and how it changed their lives. This was useful in figuring out how much harm the product had caused them emotionally.

Smith v. Department of Child Services (2019): Emotion detection tech was used to rule on emotional fitness of a child in a custody battle. During interviews and other interactions, the forensic psychologist learned the dynamics of the child's thinking and the child's interests and emotional state. Mind you, this information was essential to help create a custody plan that put the child's emotional well-being first.

Several technologies and systems are proposed to identify and label one or more speakers that implement their method of recognition and of operation, and are classified according to approach and test method used in the extraction, selection and classification of functions.

The testing not only served to quantify the emotional distress, but also illustrated its devastating effect on the plaintiffs' lives. This technology had a direct impact on the legal process, establishing a quantitative basis for judging the validity of the plaintiffs' demands and supporting the court in taking informed decisions as to compensation and liability. The case also represented how technology was being used more and more in the world of law, and especially when it came to complex emotional states and their legal implications.

3. PROPOSED WORK

In this paper, we integrated the proposed method into the feature extraction, selection and classification of the application. The experiment was performed over the three standard benchmark datasets with good results. Figure 1 represents the flow chart of Speech Emotion Recognition (SER) model using an MLP classifier.

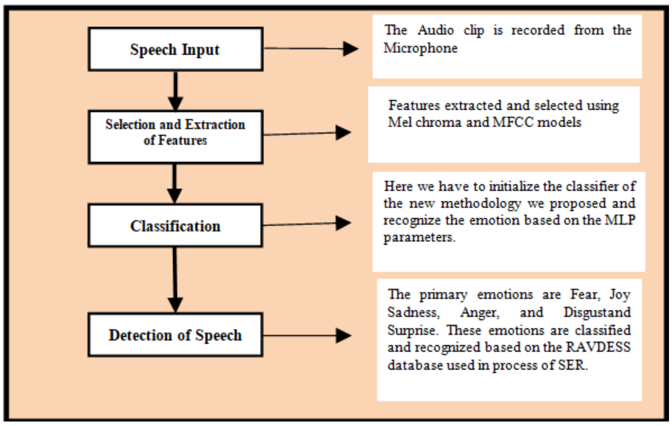

Speech Input
The Audio clip is recorded from the Microphone
Selection and Extraction of Features
Features extracted and selected using Mel chroma and MFCC models
Classification
Here we have to initialize the classifier of the new methodology we proposed and recognize the emotion based on the MLP parameters.
Detection of Speech
The primary emotions are Fear, Joy Sadness, Anger, and Disgustand Surprise. These emotions are classified and recognized based on the RAVDESS database used in process of SER.

Figure 1. Flowchart of speech emotion recognition system and multi-layer perceptron classifier

Algorithm 1 represents the step-by-step explanation of proposed model.

Algorithm 1. Speech emotion recognition algorithm

STEP 1: Microphone-assisted Speech Input: The speech signal $x(t)$, where t represents time duration of the signal. $S_i(t)$ is the original speech component, and $n(t)$ is noise.

$$x(t) = \sum_{i=1}^{N} S_i(t) + n(t)$$

STEP 2: Feature Extraction using MFCC and Mel-Frequency Methods: The signal is segmented into frames of equal length, typically using a sliding window approach:

$$x_i(t) = x(t) \cdot w(t - iT)$$

where, w(t) is a windowing function, T is the window step size, and i is the frame index. FFT is used to convert time-domain signal into frequency domain:

$$X_i(f) = F\{x_i(t)\}$$

where, F denotes the Fourier transform and $X_i(f)$ represents the frequency components.

Mel-frequency cepstral coefficients (MFCC) extraction: The frequency domain signal is filtered using a set of Mel filters, where each filter's response is triangular and spaced according to the Mel scale:

$$M_i(m) = \sum_f H_m(f)|X_i(f)|^2$$

where, $H_m(f)$ is the triangular filter and $M_i(m)$ represents the Mel-filtered output. Apply logarithm to the filter bank outputs log $M_i(m)$. Take the Discrete Cosine Transform (DCT) to obtain the MFCC features:

$$MFCC_i = D(\log\ M_i(m)).$$

where, D is the DCT, producing a feature vector of coefficients representing each frame.

STEP 3: Deep Neural Network-Based Classifier: Each frame's MFCC features form a vector:

$$F_i = [MFCC_{i1}, MFCC_{i2}, MFCC_{i3}, \dots \dots \dots, MFCC_{id}]$$

Let the input layer to the MLP have d-dimensional input (the MFCC features). Each hidden layer in the MLP transforms the input using weights $W^{(l)}$ and bias $b^{(l)}$ & used ReLU as a activation function:

$$h^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right)$$

where, l depicts layer number, $h^{(l)}$ is the hidden unit output, $W^{(l)}$ represents weights, $b^{(l)}$ depicts bias & σ(·) represents activation function:

$$\sigma(\mathrm{x}) = \max(0, x)$$

Output Layer: The final layer produces probabilities for each emotion class c (e.g., anger, sadness, joy, etc.) using the softmax function:

$$\hat{y}_c = \frac{e^{ZC}}{\sum_j e^{Zj}}$$

where, Z_c is the output of the final layer for class c, and $\hat{y}_c$ is the predicted probability for emotion class c.

Loss Function: The classifier is trained by minimizing the cross-entropy loss:

$$L = -\sum_c y_c \log\ \hat{y}_c$$

STEP 4: Emotion Recognition from Speech Input
After training, the classifier is able to predict the emotion for a given speech input. The model will predict the emotion C^* corresponding to the class with the highest predicted probability $C^* = \mathrm{argmax}_c \hat{y}_c$ and the model is trained to recognize basic emotions, which form the classification categories:

$$C^* = \{\text{anger, sadness, joy, disgust, fear}\}$$

Speech Input: The proposed work has been considered the speech as input for the further process and this standard dataset has been taken from RAVDESS etc.

Feature Extraction & Selection: These emotions separable because of the change in pitch, rate of speech, energy and delta frequency. The key features that represent emotions are computed through energy analysis, mainly short-term energy and short-term average energy.

In this work, we have considered the features such as: pitch, energy, Linear LPC Coding, Mel-frequency spectrum coefficients (MFCCs) and utilized in extracting and selecting of features for SER. The extraction and selection function are illustrated as shown in Figure 2.

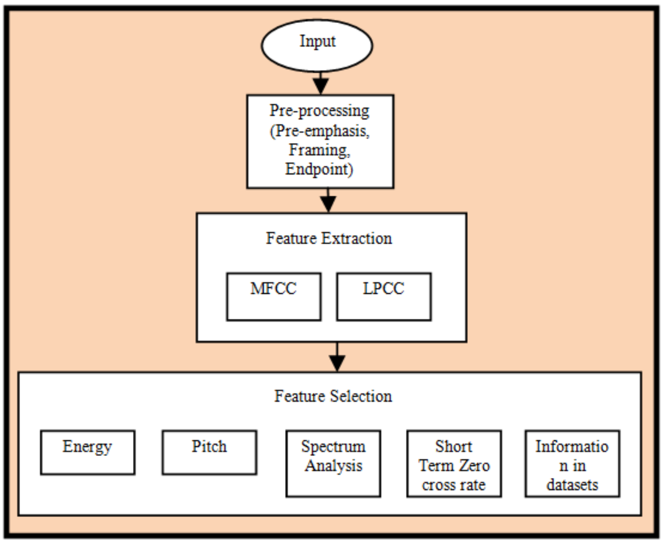


Figure 2. Feature extraction and selection based on the input feature

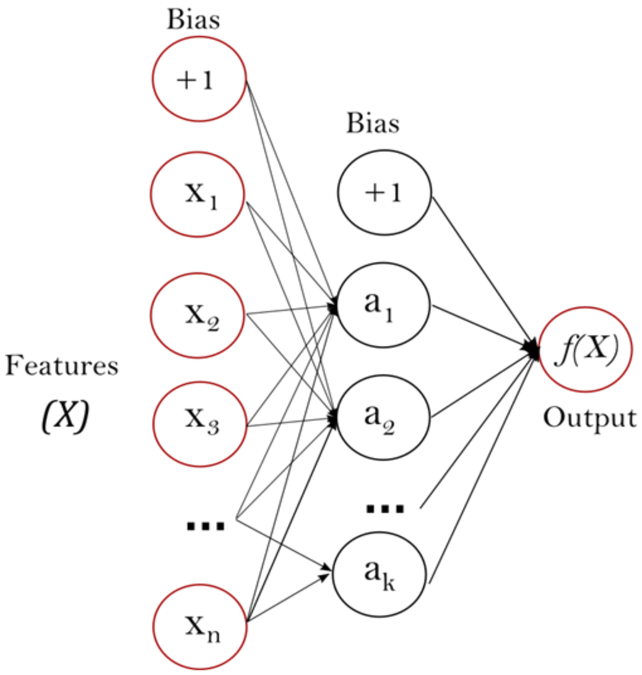


Figure 3. Artificial Neural Network (ANN)-based MLP classifier with a single hidden layer

Classification Model: Several classifiers have been used for the task of SER. A data centric approach discovers the best outcomes for SER and then analyze SER systems on various standard datasets. Some authors work with commercial databases, whereas others develop their own databases.

The proposed approach is projected to optimized and the recommended SER output by using the best speech features. The classification method relies on cognitive intuition of the grades and is able to identify real feelings mapping.

The challenge here is to derive suitable classification for SER. The multilayer perceptron (MLP) classification is the type of the ANN, which is mainly used for speech emotion recognition. MLP has three layers of nodes – input, hidden, and output layers as shown in Figure 3.

The MLP works under a supervised back propagation learning paradigm and MLP involve multiple layers of nonlinear activation. Based on multiple layers of nonlinear activation functions emotions were taken for classification process. First, the voice is turned into text, then the text is changed into binary form (0s and 1s) using a translation method. Finally, binary data is given to the neural network, where the classifier creates synaptic weights to form neural features. Each hidden layer neuron receives outputs from all neuron in the previous layers, and through a weighted linear sum, it gives the output value, where n is the nth layer of neurons, and j is one of the weight vector components. The output layer takes the values from the last hidden layer and calculates a linear sum. The hidden layers are influenced by the input signals. In our model, we used 17 layers to recognize emotions as shown in Figure 4.

The final emotion results are obtained by the mathematical modelling as describe below. A Perceptron produces a single output from multiple inputs by combining them into a linear equation, sometimes followed by a nonlinear activation function as shown in Eq. (1).

$$y = \varphi(\textstyle\sum_{i=1}^{n} \omega_i x_i + b) = \varphi(W^T X + b) \quad (1)$$

Above, x is the input, b is a bias and u is a nonlinear activation function, w is a weight vector. A model is a set of parameters (weights and biases) that uses the input with minimum errors from the desired output during the time of the training. Finally, an illustrative procedure can be applied to access the error and return is used to update the weight and bias with respect to it as shown in Eq. (2).

$$W \to W - \eta = \alpha \frac{\partial R(W)}{\partial W} + \frac{\partial Loss}{\partial W} \quad (2)$$

Encoding the range (106) allows the MLP reduce error by one step. The perceptron uses weights, an addition unit, and a damping threshold, where a numeron represents a weighted input or output 1 if the sum > some parameterized threshold score 0 otherwise as shown in Eqs. (3) and (4).

$$w_1 x_1 + w_2 x_2 \,....+ w_n x_n > 0 \; then\ output\ is\ 1 \quad (3)$$

$$w_1 x_1 + w_2 x_2 \,....+ w_n x_n \leq 0 \; then\ output\ is\ 0 \quad (4)$$

The input and connection weights are generally assumed to be real. If the estimated output is the same as the desired one, and this result is found to be acceptable and thus no updates are done on the weights, further input will be given to perceptron and weights will adjust in such a way that it can minimize the errors. But the output is not similar to the desired output. So, weights are adjusted again as shown in Eq. (5):

$$\Delta W = \eta \times d \times X \quad (5)$$

where, X is input data, ΔW is predicted output and d represent learning rate.

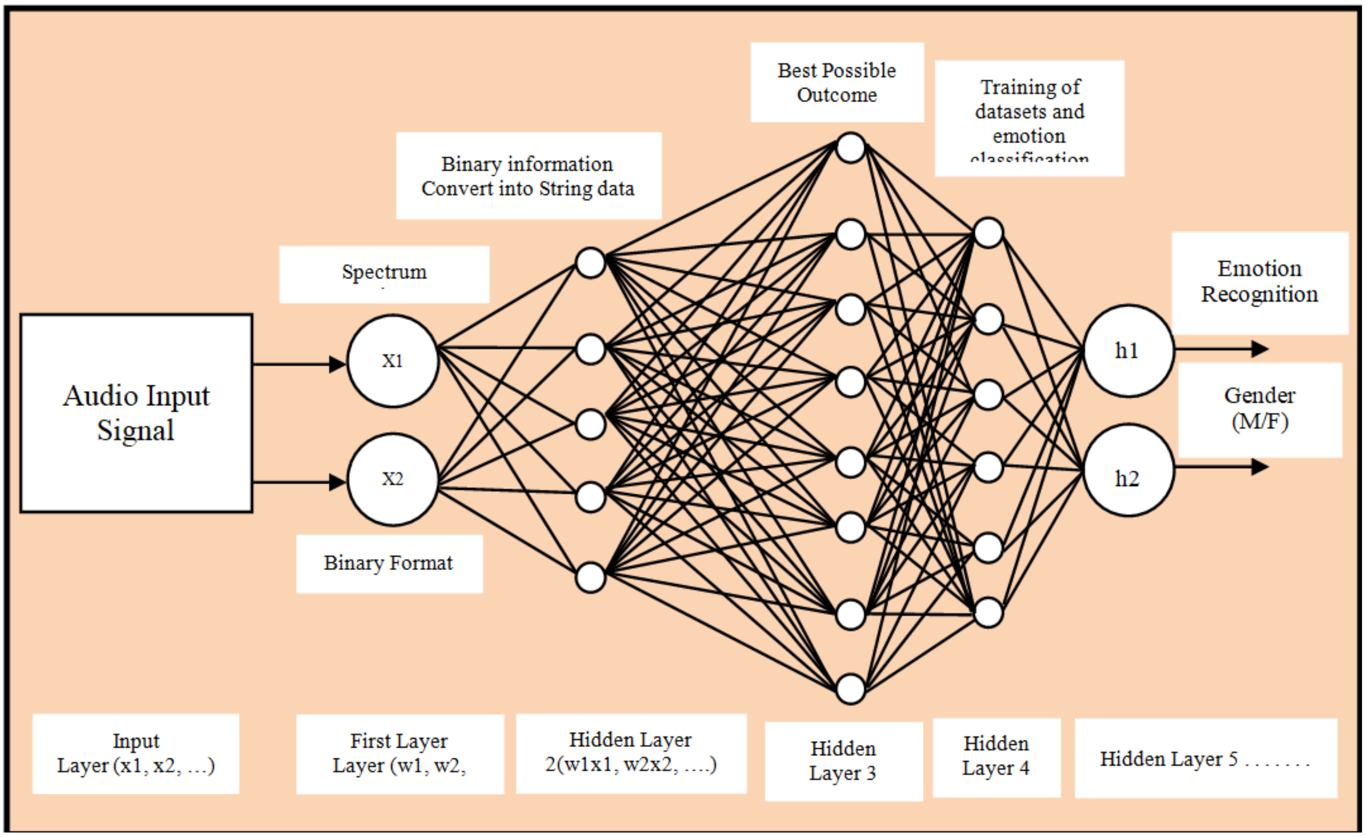
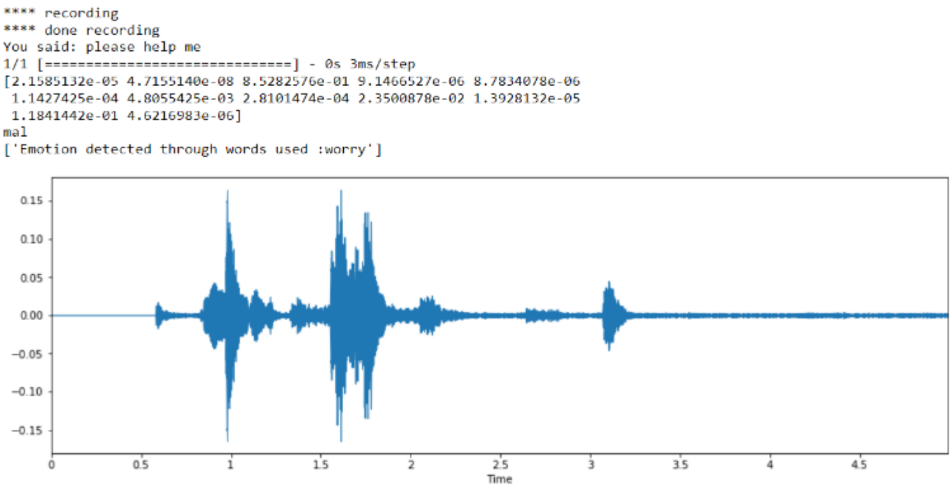
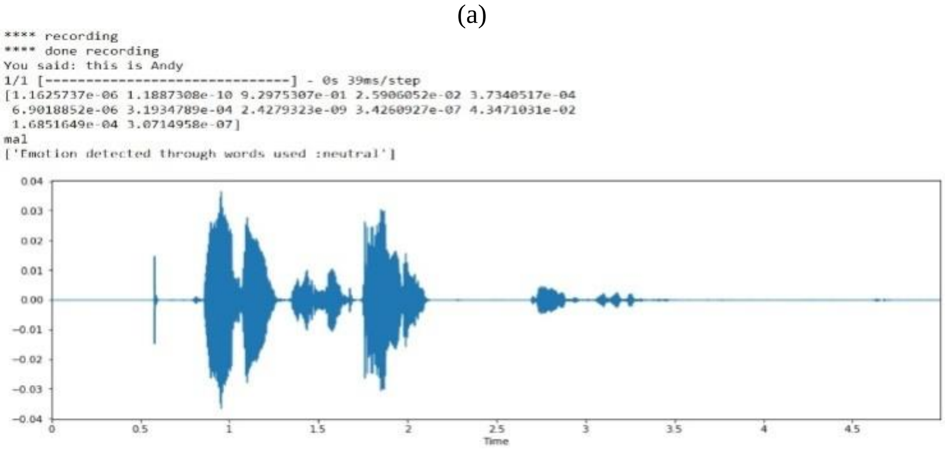

Audio Input Signal
Spectrum
X1
X2
Binary Format
Binary information Convert into String data
Best Possible Outcome
Training of datasets and emotion classification
h1
h2
Emotion Recognition
Gender (M/F)
Input Layer (x1, x2, ...)
First Layer Layer (w1, w2,
Hidden Layer 2(w1x1, w2x2, ...)
Hidden Layer 3
Hidden Layer 4
Hidden Layer 5 . . . .

Figure 4. Process of MLP classifier based on ANN



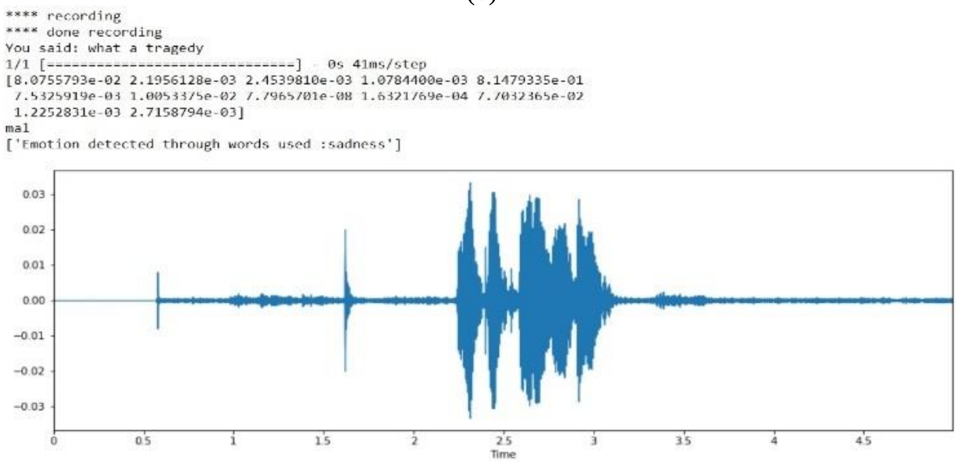
```
**** recording
**** done recording
You said: please help me
1/1 [==============================] - 0s 3ms/step
[2.1585132e-05 4.7155140e-08 8.5282576e-01 9.1466527e-06 8.7834078e-06
 1.1427425e-04 4.8055425e-03 2.8101474e-04 2.3500878e-02 1.3928132e-05
 1.1841442e-01 4.6216983e-06]
mal
['Emotion detected through words used :worry']
```

(a)



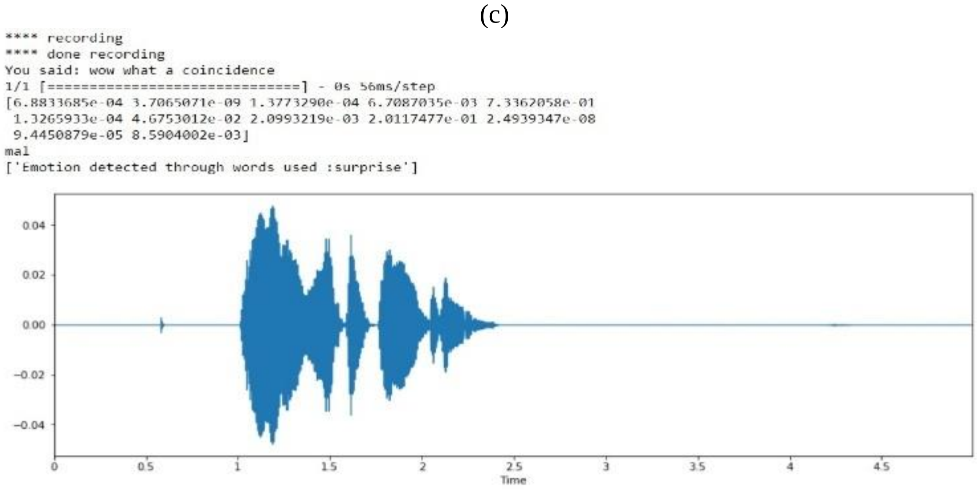
```
**** recording
**** done recording
You said: this is Andy
1/1 [------------------------------] - 0s 39ms/step
[1.1625737e-06 1.1887308e-10 9.2975307e-01 2.5906052e-02 3.7340517e-04
 6.9018852e-06 3.1934789e-04 2.4279323e-09 3.4260927e-07 4.3471031e-02
 1.6851649e-04 3.0714958e-07]
mal
['Emotion detected through words used :neutral']
```

(b)



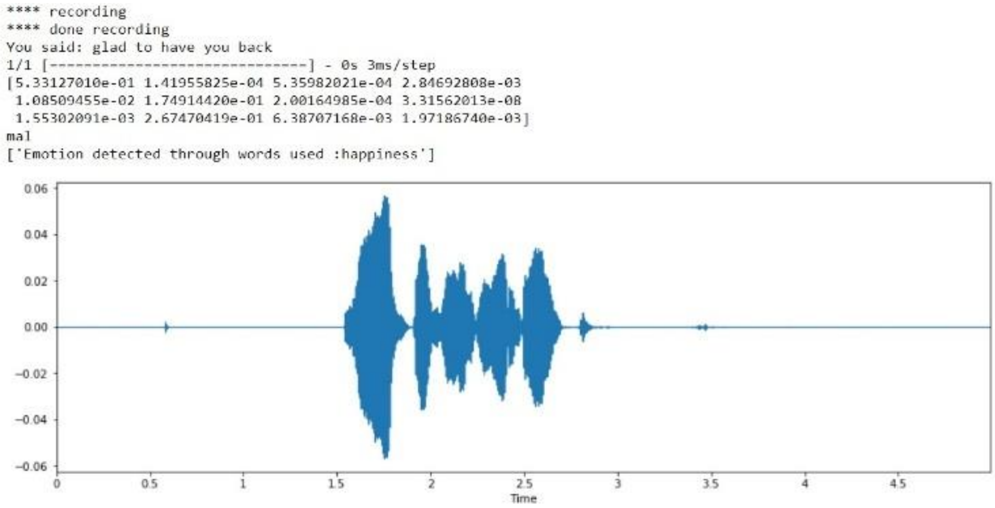
```
**** recording
**** done recording
You said: what a tragedy
1/1 [==============================] - 0s 41ms/step
[8.0755793e-02 2.1956128e-03 2.4539810e-03 1.0784400e-03 8.1479335e-01
 7.5325919e-03 1.0053375e-02 7.7965701e-08 1.6321769e-04 7.7032365e-02
 1.2252831e-03 2.7158794e-03]
mal
['Emotion detected through words used :sadness']
```

(c)



```
**** recording
**** done recording
You said: wow what a coincidence
1/1 [==============================] - 0s 56ms/step
[6.8833685e-04 3.7065071e-09 1.3773290e-04 6.7087035e-03 7.3362058e-01
 1.3265933e-04 4.6753012e-02 2.0993219e-03 2.0117477e-01 2.4939347e-08
 9.4450879e-05 8.5904002e-03]
mal
['Emotion detected through words used :surprise']
```

(d)



```
**** recording
**** done recording
You said: glad to have you back
1/1 [==============================] - 0s 3ms/step
[5.33127010e-01 1.41955825e-04 5.35982021e-04 2.84692808e-03
 1.08509455e-02 1.74914420e-01 2.00164985e-04 3.31562013e-08
 1.55302091e-03 2.67470419e-01 6.38707168e-03 1.97186740e-03]
mal
['Emotion detected through words used :happiness']
```

(e)

Figure 5. (a)-(e) Recognition outputs of emotion recognition on audio signals
Note: X-axis: Time period of the audio signal recording; Y-axis: Frequency of the speech signal in decibels (dB).

Recognized Emotion: The emotions are classified using ANN architecture. There are six basic emotions as shown in Figure 5 (a)-(e) and it has to be categorized, and their database will be utilized by the SER. Speak rate, power and spectrum are the features that align with the appearances of the emotions below. Emotion Recognition is difficult due to different contexts, cultures, facial reaction to individuals, and equivocal literature.

Deep Learning [3] is designed to perform tasks by learning from data as shown in Figure 6 and deep neural networks (DNNs) are used in areas like image recognition, classification, decision-making, and pattern detection [4]. Other methods, such as multimodal deep learning, are also used to improve image and speech recognition tasks [5]. In our case, we train a DNN model to predict the likelihood of each emotion for every segment based on segment-level features. The DNN acts as a mood detector for each part and the patterns in these segments can still be analyzed and used in a higher-level model [18] to predict the emotion of the full utterance as shown in Eq. (6).

$$t = [P(E_1), \dots . P(E_K)\,]^T \tag{6}$$

The DNN input nodes match the segment vector's dimension. It concentrates on the output layer with a size = K Part B: Cross validation will be performed on the number of the hidden layers and the units of the hidden layers. As an objective, the trained DNN must produce the probability distribution (t) over all emotional states for each segment. Deep learning training uses basic psychological models, where each model helps estimate emotions. The common emotions are anger, fear, happy, sadness and neutral. Dimensional models depict specific emotions i.e. valence and arousal. The training of DNN happens on word embedding, which are the main features. Word Embedding is the named category of taking words or phrases from vocabulary and returning their corresponding number vectors in order to convert to a sequence of language models and NLP features.

Then in continuous space, we need to encode words that measure semi-semantic distances between the words. The syntactic identity, which pressures the sentences in which semantics are order-sensitive the on readings, is captured. We have to learn the syntax in here. Instead, it was learned automatically, how to recognize emotions on deep learning based on word embeddings quantified semantics in data. (The DNN model catching syntactic features on RNNs. This forms the basis of the framework for deep learning, which is trained by DNN, and the emotions are classified using the web-embedding method in shown in the Figure 6. The five states in the study are (happiness, neutral, excitement, frustration, sadness), and there are five states each of them has their corresponding dimensions in the DNN [17-23].

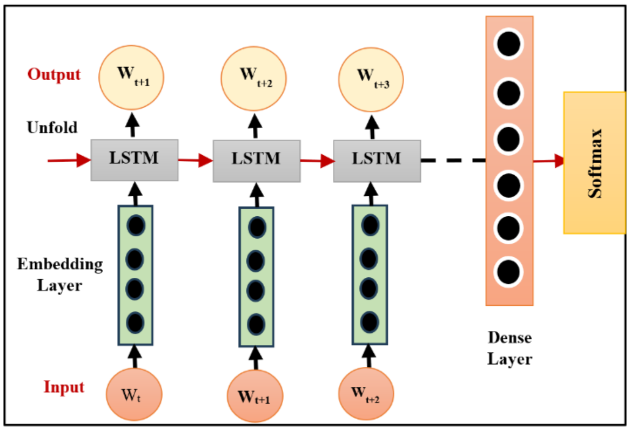


Figure 6. Deep learning architecture

4. RESULT AND ANALYSIS

Entire testing was carried out using system configuration on Python 3.8, 3 benchmark datasets i.e. RAVDNESS, TIMIT Corpus and Emo-DB The aspects which monitored the changes have been underlined in bullet form: Ram 8GB Intel Core I5 RADEON Graphics Card etc.

4.1 Standard database

Here we employ RAVDESS having 7356 files. The Collection features 24 male and female contributors, speaking two words in a neutral North American accent. Speech encompasses calm, happy, sad, angry, fear, surprise, and disgust expressions, and song includes calm, happy, sad, angry, and fear expressions. Two intensities (normal/strong) are created for each expression, together with one neutral expression. The emotional reports are provided along with the database mark-tab and classifies as different .wav files into seven emotions as discussed above. Servlet Exception in emotion classification for 4 sec. Speech signals are transformed by the emotional class in the. wav format. TIMIT and EMO-DB are also employed in other databases.

4.2 Comparison of the datasets

The datasets are compared based on emotions, form, granularities, emotions, size and description in the Table 1.

Table 1. Description of datasets and specifications

Dataset	Emotions	Description
EMO-DB	6 Emotions	Features contain around 500 words delivered by actors in an angry, frightning, fearful, bored and gender-inclusive manner.
TIMIT	7 Emotions	TIMIT consisted of 630 broadband recordings of 10 phonetically rich sentences for each of the eight major American dialects.
RAVDESS	7 Emotions	Target stimuli were recorded in a sound-damped studio from 24 actors (12 female, 12 male) who spoke two lexically-balanced sentences in a neutral North American accent.

4.3 Overall result

The emotions of the voice signal are detected and recognized based on different criteria. The general evaluation for performance can be inferred on the basis of the SER and its performance efficiency which are compared to other databases in the aforementioned works. Tables 2-5 and Figures 7-10 are representing the proposed work outcomes in term of accuracy, time and error and learning rate on all three standard datasets.

The proposed approach based on MLP algorithm outperforms SVM, CNN, DNN, and ensemble techniques on all three tested datasets, Emo-DB, TIMIT Corpus, and RAVDESS. We reached the best accuracy of 86.20% with the ideal error rate of 36.0% in 0.01s – among all models – for the Emo-DB database, and yet kept computation time close to other models. It achieved at the best the accuracy of 87.60% for TIMIT Corpus, which is surpassed all systems: Bagged Ensemble of SVMs 1(84.12%) and DNN+Decision Tree SVM (82.33%). The introduced model also achieved less error rate (51.30%) and less computational time (2.18s) with respect to many of the previous methods. Also, the MLP-based ANN obtained the best classification accuracy of 80.21% on RAVDESS dataset with a significantly low error rate (47.80%) and computational time (5.61s) compared to rest of the models. These findings (shown in Table 6) in aggregate demonstrate the effectiveness, efficiency, and reliability of the proposed approach on various datasets.

Table 2. Results of SER as per SPAB score in (%)

Dataset	WER	AA	CER	RR	SPL	SNR	SPAB	Accuracy (%)
EMO-DB	0.9712	1	0.967	0.89	0.951	0.961	5.201	**86.2**
TIMIT	0.9371	1	0.89	0.76	0.921	0.945	5.104	**85.43**
RAVDESS	0.9233	0.941	0.9	0.71	0.869	0.881	4.712	**80.21**

Table 3. SER results based on confusion matrix (%)

Emotion Category	Anger	Happiness	Worry	Neutral	Surprise	Hate
Anger	**76**	7.90	1.00	1.6	11.0	1.50
Happiness	7	**80.50**	2.50	3	14.80	2.20
Worry	2.5	2.50	**77.0**	3	1.50	34.20
Neutral	0.7	2.30	2.50	**92**	1.50	1.0
Surprise	11.5	8.00	3.00	1	**86.50**	0
Hate	1.5	2.50	28.0	3	0	**65.0**

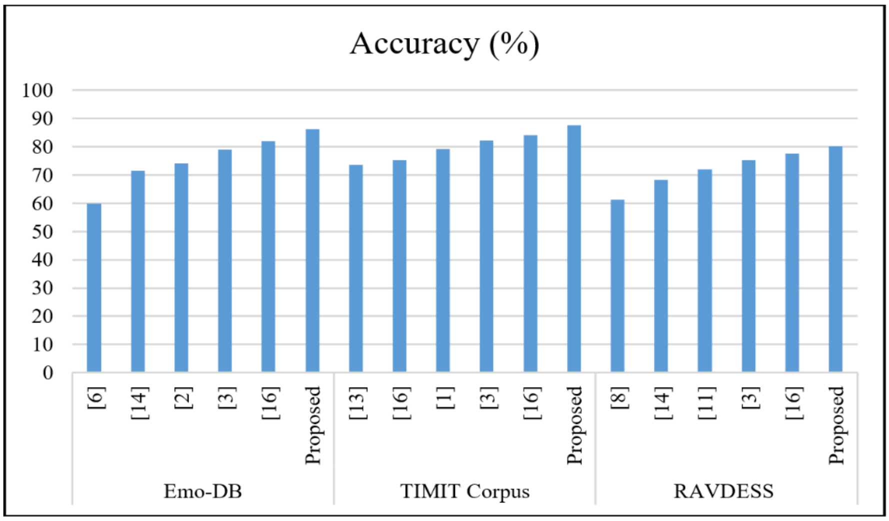


Figure 7. Accuracy comparison of proposed work with existing methods on EMO, TIMIT and RAVDESS database

Table 4. Average recognition and learning rate of proposed work across various layer

Layers	1st Layer	2nd Layer	3rd Layer	4th Layer	5th Layer
ARR	70.19	71.19	71.06	59.1	66.76
LR	0.001	0.005	0.003	0.008	0.1

Note: Avg Recognition Rate: ARR; Learning Rate: LR.

Table 5. Average recognition rate of proposed algorithm

Emotion Category	Anger	Happiness	Worry	Neutral	Surprise	Hate
RR	76	80.5	77	92	86.5	64
ARR	80.21					

Note: Recognition Rate: RR; Avg Recognition Rate: ARR.

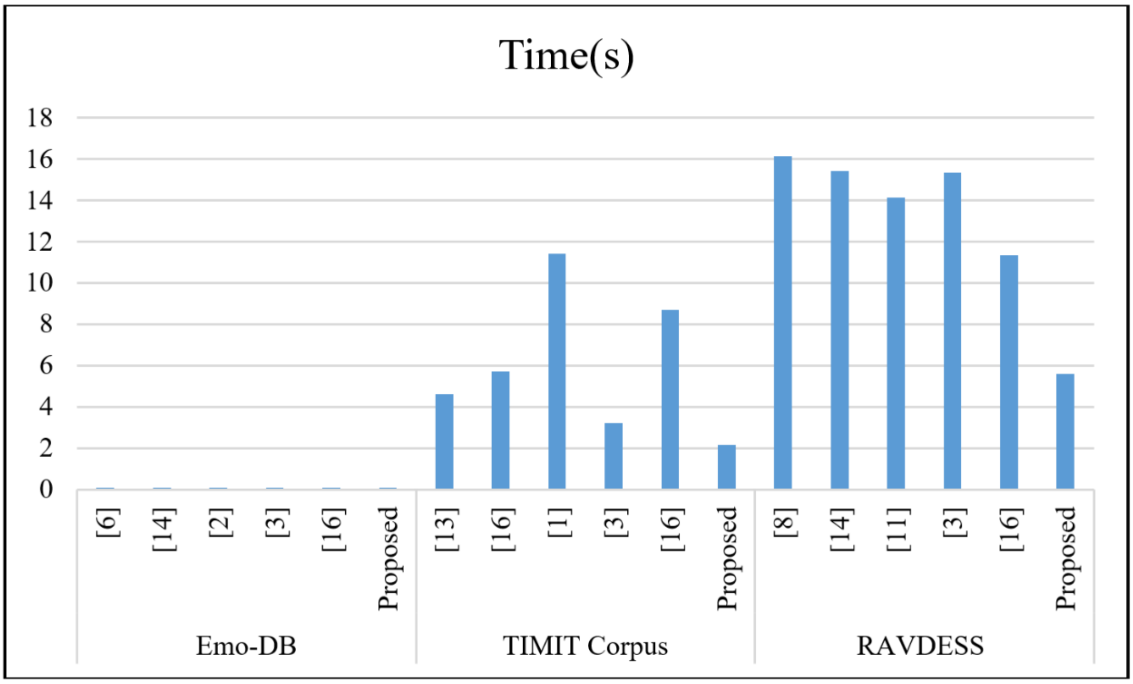


Figure 8. Time comparison of proposed work with existing methods on EMO, TIMIT and RAVDESS database

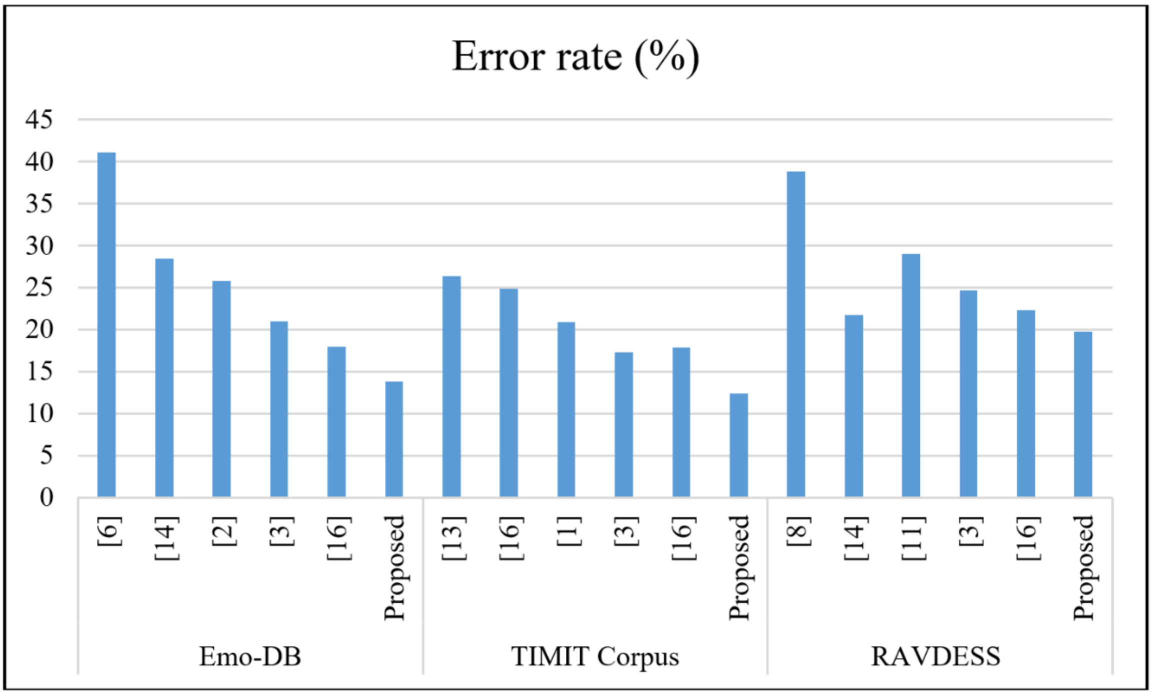


Figure 9. Comparative analysis of error rates

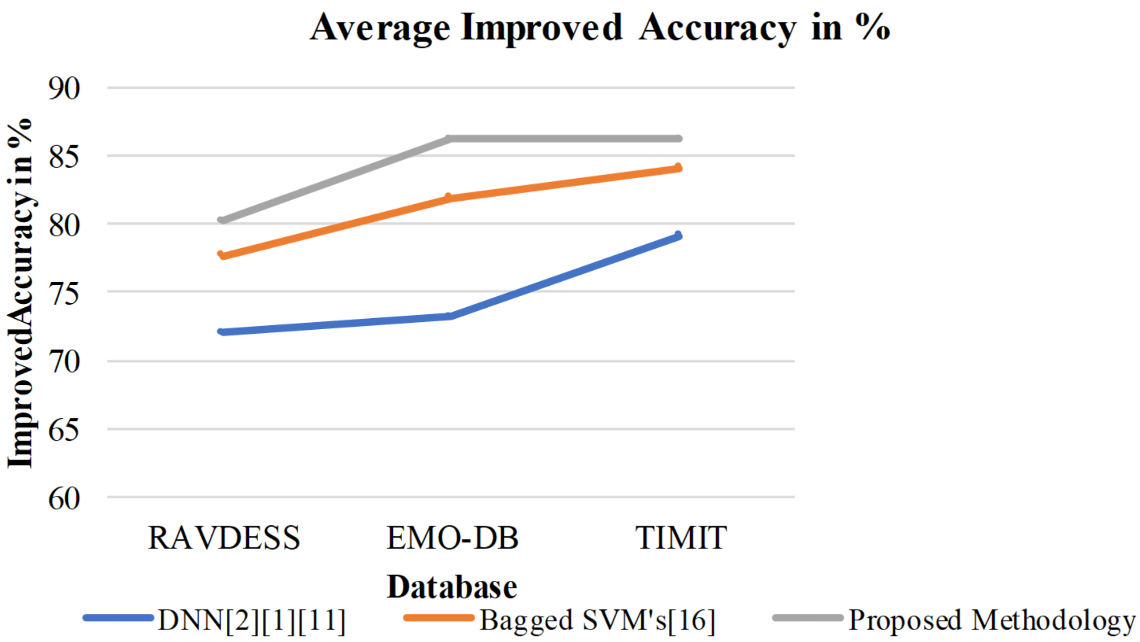


Figure 10. Average accuracy comparison with different database

Table 6. Comparative analysis of the state of art methods & proposed work

Dataset	Reference	Accuracy (%)	Time(s)	Error Rate (%)	ANOVA F-Value	p-Value	GPU Hours	Model Size (Parameters)
Emo-DB	[6]	59.90	0.01	41.10			2	50M
	[14]	71.59	0.01	28.41			3	55M
	[2]	74.19	0.01	25.81			4	58M
	[3]	79.01	0.01	20.99			5	60M
	[16]	82.02	0.01	17.98			6	65M
	Proposed	**86.20**	**0.01**	**13.80**	**10.47**	**0.001**	7	70M
TIMIT Corpus	[13]	73.67	4.61	26.33			10	45M
	[16]	75.18	5.71	24.82			12	50M
	[1]	79.12	11.41	20.88			13	55M
	[3]	82.33	3.21	17.33			8	57M
	[16]	84.12	8.71	17.82			14	60M
	Proposed	**87.60**	**2.18**	**12.40**	**9.82**	**0.002**	9	65M
RAVDESS	[8]	61.20	16.12	38.80			12	40M
	[14]	68.30	15.43	21.70			13	45M
	[11]	72.01	14.12	28.99			11	50M
	[3]	75.32	15.35	24.68			15	55M
	[16]	77.69	11.32	22.31			16	60M
	Proposed	**80.21**	**5.61**	**19.79**	**8.94**	**0.003**	10	65M

5. CONCLUSIONS

Proposed work provides a reliable and objective methodology to measure emotional states, and is of high importance for both mental health assessments and forensic psychology. Based on emotional information carried by behaviour in images, the system supports assessments of credibility of witnesses and detection of fraud or trauma, thus contributing to the objectivity for the assessment of forensically relevant endpoints and legal procedures. Experimental results on benchmark speech emotion databases show that the proposed method achieves superior performance over state-of-the-art algorithms and methods in terms of recognition rate, accuracy, error rate, and speed. The high degree of accuracy obtained in different emotional groups confirms the practical performance of the model. In addition, its use for mental health diagnostics overcomes the shortcomings of conventional subjective evaluation and provides a systematic and data-driven approach to characterize emotional states. The neural network-based design, tested on extensive cross-validation over different datasets, provides a robust framework for the development of emotion recognition technologies in practical application in healthcare and law gaining power.

REFERENCES

[1] Basharirad, B., Moradhaseli, M. (2017). Speech emotion recognition methods: A literature review. AIP conference proceedings, 1891(1): 020105. https://doi.org/10.1063/1.5005438

[2] Özseven, T. (2019). A novel feature selection method for speech emotion recognition. Applied Acoustics, 146: 320-326. https://doi.org/10.1016/j.apacoust.2018.11.028

[3] Sun, L., Zou, B., Fu, S., Chen, J., Wang, F. (2019).

Speech emotion recognition based on DNN-decision tree SVM model. Speech Communication, 115: 29-37. https://doi.org/10.1016/j.specom.2019.10.004

[4] Khanchandani, K.B., Hussain, M.A. (2009). Emotion recognition using multilayer perceptron and generalized feed forward neural network. Journal of scientific and industrial research, 68(5): 367-371.

[5] Kerkeni, L., Serrestou, Y., Mbarki, M., Raoof, K., Mahjoub, M.A., Cleder, C. (2019). Automatic speech emotion recognition using machine learning. In Social Media and Machine Learning, pp. 1-16.

[6] Ke, X., Zhu, Y., Wen, L., Zhang, W. (2018). Speech emotion recognition based on SVM and ANN. International Journal of Machine Learning and Computing, 8(3): 198-202.

[7] Albornoz, E.M., Milone, D.H., Rufiner, H.L. (2011). Spoken emotion recognition using hierarchical classifiers. Computer Speech & Language, 25(3): 556-570. https://doi.org/10.1016/j.csl.2010.10.001

[8] Cen, L., Wu, F., Yu, Z.L., Hu, F. (2016). A real-time speech emotion recognition system and its application in online learning. In Emotions, Technology, Design, and Learning, pp. 27-46. https://doi.org/10.1016/B978-0-12-801856-9.00002-5

[9] Aouani, H., Ayed, Y.B. (2020). Speech emotion recognition with deep learning. Procedia Computer Science, 176: 251-260. https://doi.org/10.1016/j.procs.2020.08.027

[10] Tyagi, R., Agarwal, A. (2018). Emotion detection using speech analysis. Science, 3(3): 18-20.

[11] Chen, L., Mao, X., Xue, Y., Cheng, L.L. (2012). Speech emotion recognition: Features and classification models. Digital Signal Processing, 22(6): 1154-1160. https://doi.org/10.1016/j.dsp.2012.05.007

[12] Rázuri, J.G., Sundgren, D., Rahmani, R., Moran, A., Bonet, I., Larsson, A. (2015). Speech emotion recognition in emotional feedback for human-robot interaction. International Journal of Advanced Research in Artificial Intelligence (IJARAI), 4(2): 20-27. https://doi.org/10.14569/IJARAI.2015.040204

[13] Milton, A., Selvi, S.T. (2014). Class-specific multiple classifiers scheme to recognize emotions from speech signals. Computer Speech & Language, 28(3): 727-742. https://doi.org/10.1016/j.csl.2013.08.004

[14] Parthasarathy, S., Tashev, I. (2018). Convolutional neural network techniques for speech emotion recognition. In 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), Tokyo, Japan, pp. 121-125. https://doi.org/10.1109/IWAENC.2018.8521333

[15] Idris, I., Salam, M.S.H., Sunar, M.S. (2016). Speech emotion classification using SVM and MLP on prosodic and voice quality features. Jurnal Teknologi (Sciences & Engineering), 78(2-2): 27-33. https://doi.org/10.11113/jt.v78.6925

[16] Bhavan, A., Chauhan, P., Shah, R.R. (2019). Bagged support vector machines for emotion recognition from speech. Knowledge-Based Systems, 184: 104886. https://doi.org/10.1016/j.knosys.2019.104886

[17] Watson, J., Aglionby, G., March, S. (2023). Using machine learning to create a repository of judgments concerning a new practice area: A case study in animal protection law. Artificial Intelligence and Law, 31(2): 293-324. https://doi.org/10.1007/s10506-022-09313-y

[18] Kumar, S., Rani, S., Jain, A., Verma, C., Raboaca, M.S., Illés, Z., Neagu, B.C. (2022). Face spoofing, age, gender and facial expression recognition using advance neural network architecture-based biometric system. Sensors, 22(14): 5160. https://doi.org/10.3390/s22145160

[19] Kumar, S., Jain, A., Rani, S., Alshazly, H., Idris, S.A., Bourouis, S. (2022). Deep neural network based vehicle detection and classification of aerial images. Intelligent Automation & Soft Computing, 34(1): 119-131. ttps://doi.org/10.32604/iasc.2022.024812

[20] Rajawat, S. (1972). R.M. Malkani v. State of Maharashtra. Naya Legal. https://www.nayalegal.com/rm-malkani-v-state-of-maharashtra.

[21] Tyagi, A., Mohan, M. (2021). Distinction between the direct evidence and circumstantial evidence. Available at SSRN 3838902. https://doi.org/10.2139/ssrn.3838902

[22] Laddhad, V. (2023). An examination of the dynamics and obstacles facing the legal importance of dying declarations in existing criminal jurisprudence. Jus Corpus Law Journal, 4: 176.

[23] Gonzalez, E.A., Teal, K.B. (2015). No ideas but in things: A practitioner's look at demonstrative evidence. The Florida Bar Journal, 89: 10-16.

[24] Heiser, W.W. (1996). A critical review of the local rules of the United States District Court for the Southern District of California. San Diego Law Review, 33: 555.

[25] Schmidt, J. (2013). School desegregation in New Castle County, Delaware. Metropolitan Desegregation, 37.

[26] Park, S., James, J.I. (2024). Lessons learned building a legal inference dataset. Artificial Intelligence and Law, 32(4): 1011-1044. https://doi.org/10.1007/s10506-023-09370-x

[27] Vianna, D., de Moura, E.S., da Silva, A.S. (2024). A topic discovery approach for unsupervised organization of legal document collections. Artificial Intelligence and Law, 32(4): 1045-1074. https://doi.org/10.1007/s10506-023-09371-w

[28] Francesconi, E., Governatori, G. (2023). Patterns for legal compliance checking in a decidable framework of linked open data. Artificial Intelligence and Law, 31(3): 445-464. https://doi.org/10.1007/s10506-022-09317-8

[29] Watson, J., Aglionby, G., March, S. (2023). Using machine learning to create a repository of judgments concerning a new practice area: A case study in animal protection law. Artificial Intelligence and Law, 31(2): 293-324. https://doi.org/10.1007/s10506-022-09313-y