Speech Emotion Recognition: Enhancing Human-Computer Interaction

Marwan Salah
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
marawan20191700618@cis.asu.edu.eg

Alaa Adel
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
alaa20191700388@cis.asu.edu.eg

Abdelaziz Gamal
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
abdelaziz20191700376@cis.asu.edu.eg

Abdelrahman Alaa
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
abdelrahman20191700365@cis.asu.edu.eg

Abdelrahman Amr
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
abdelrahman20191700368@cis.asu.edu.eg

Samar Aly
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
samar.Aly@cis.asu.edu.eg

Sally Saad
Computer Science Department
Faculty of Computer and Information Science Ain Shams University
Cairo, Egypt
sallysaad@cis.asu.edu.eg

Abstract—In today's world, human-computer interactions (HCI) aim to replicate genuine human-to-human communication, wherein emotions play a pivotal role. The human voice carries a rich array of emotions, making speech a natural and powerful means of communication. Emotions greatly influence human interactions, encompassing facial expressions, body gestures, voice characteristics, intonation, and linguistic contents. This research discusses the impact of the classification approach, identifying the most appropriate combination of features and data augmentation on speech emotion detection accuracy. Selection of the correct combination of handcrafted features with the classifier plays an integral part in reducing computation complexity. Deep learning (DL) has revolutionized speech emotion recognition (SER) by replicating neural processes, learning complex patterns from raw data, and significantly improving classification accuracy. The suggested classification model, a 1-Dimensional Convolutional Neural Network (1D CNN), outperforms traditional Machine learning (ML) approaches in classification. Unlike most earlier studies, which examined emotions primarily through a single language lens, this analysis looks at numerous language data sets. With the utilization of the most discriminating features and data augmentation techniques, the SER system achieves 91% testing accuracy, 94% validation accuracy, and 97% training accuracy for the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and Toronto Emotional Speech Set (TESS) datasets combined.

Keywords: Human-Computer Interaction, Deep Learning, Emotion Detection, RAVDESS.

1. INTRODUCTION

Speech, as the most natural form of communication between humans, carries valuable information about the characteristics of the speaker including gender, emotional state, and more. With the advancement of technology and the increased prevalence of HCI, SER has emerged as a crucial area for improving human-machine communication and interaction [1]. SER aims to bridge the gap between the physical and digital worlds by equipping machines with the ability to perceive and understand emotions, ultimately enhancing the quality of human-machine interaction. Researchers focus on leveraging DL models to improve SER efficiency and effectiveness, resulting in diverse architectures excelling across multiple domains in neural networks. The types of neural network architectures include feed-forward architectures and recurrent architectures, such as Deep Neural Networks (DNNs) and CNN have proven highly effective in tasks involving image and video processing, as well as speech recognition. On the other hand, recurrent architectures, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) RNNs, have shown significant promise in SER [2]. Several features have been employed in SER, such as Mel Spectrogram, Frequency Cepstral Coefficients (MFCCs), pitch features, and auditory speech characteristics [3]. These features serve as informative representations of speech data and aid in capturing emotion-related patterns and cues. Notably, several databases, such as RAVDESS and TESS have been extensively used in these studies to facilitate the evaluation and benchmarking of SER algorithms [4]. This paper proposes a powerful DL model using vocal data to accurately recognize and interpret emotions from speech signals, bridging the gap between humans and machines. The paper highlights the significance of data augmentation and hyperparameter optimization in improving SER models' performance. Extensive experiments demonstrate the effectiveness of the proposed approach in accurately recognizing and classifying emotions from speech data, bridging the gap between humans and machines.

2. RELATED WORK

The initial exploration of this topic can be attributed to "Daellert et al. 1996" [5].Nevertheless, the concept itself predates this publication, as evidenced by the existence of a patent from the late 1970s that utilized autonomic nervous system measurements for emotion recognition.

Yazdani, et al [6] obtained a natural HCI with two steps, Feature extraction and feature classification using Sharif Emotional Speech Database (SHEMO) data set and using signal features in low- and high-level descriptions (HLDs) and different DL & ML techniques. Low level descriptions: Pitch, voicing probability, frame energy, zero crossing rates, MFCC. High level descriptions: mean, variance, min, max, median, quartiles, higher order moments.

Kanani, et al [7] proposed structure of CNN with a Comprised set of convolution, pooling and fully connected layers. He solved a multi-class classification problem where considering un- weighted mean parameter for calculating the average for this classification. The performance indicators for evaluating the models are done using a standard confusion matrix. Finally, get Accuracy 82.99% from dataset RADVESS.

Singh, et al [8] played the audio in order to hear, plotted audio features. He Extracted the characteristics. Converting one data frame and displaying structured form. Further it compares loaded models by predict function batch size thirty-two. It displays the output from the audio file what sort of expression/emotion that audio file has. After training various models it came out with the most optimum accuracy of 82% with SoftMax activation layer, “rmsprop” activation layer,18 layers, Batch-Size = 32 and with 1000 epochs.

Dong, et al [9] build two parallel CNN to extract spatial features and a transformer encoder network to extract temporal features, classifying emotions from one of 8 classes taking advantage of CNN`s advantages in spatial feature representation and sequence encoding conversion. Obtained an accuracy of 80.46% on the hold-out test set of the RAVDESS data set.

Aouani, et al [10] proposed an emotion recognition system based on speech signals in a two-stage approach, namely feature extraction and classification engine. Firstly, two sets of features are investigated which are: the first one, researcher extracted a 42-dimensional vector of audio features including thirty-nine coefficients of MFCC, Zero Crossing Rate (ZCR), Harmonic to Noise Rate (HNR) and Teager Energy Operator (TEO). The second one proposed the use of the method auto-encoder for the selection of pertinent parameters from the parameters previously extracted. Secondly, Support Vector Machine (SVM) was used as a classifier method. Experiments are conducted on the Ryerson Multimedia Laboratory (RML).

E Yu Shchetinin, et al [11] investigated the architecture of deep neural networks for recognizing human emotions from speech. CNN and RNN with a LSTM memory cell were used as models of DNN. An ensemble of neural networks was also built on their basis. Computer experiments on the use of the proposed DL models and basic ML algorithms for recognizing emotions in human speech.

Xiangmin Lun, et al [12] Proposed 64 statistical features of the speech signal including short-term energy, pitch, frame, format, and spectrum energy were extracted with speech emotion database. Mean Impact Value (MIV) and the improved Correlation-based Feature Selection (CFS) were employed to select the most influential feature set. Back Propagation Neural Network (BPNN) was used to identify the accuracy. The proposed MIV-CFS method selected the features related to speech emotion, with less recognition error.

Althaf Hussain Basha, et al [13] improved the accuracy of the speech emotion prediction using DL models. his work experiments with the Multi-Layer Perceptron (MLP) and CNN classification models on three benchmark datasets with 5700 speech files of seven emotions categories. The proposed model showed improved accuracy.

3. METHODOLOGY

The research focuses on developing a real-time SER system that accurately classifies emotions from input audio signals. The system uses preprocessing techniques, feature extraction, model classification, and view emotion detection to analyze and classify signals. The SER system operates continuously until the user stops recording, allowing users to record and view their emotions. As shown in **Figure1**.

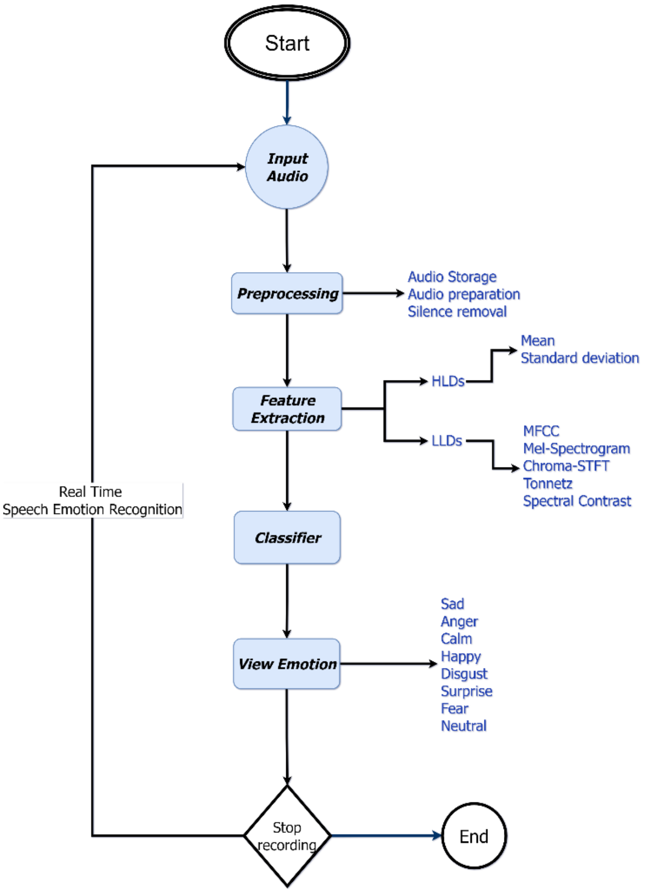

Figure 1. Real time SER block diagram

3.1 Data Collection

The databases used in the SER system, including RAVDESS and TESS, were employed to provide a diverse range of emotional speech samples suitable for training and evaluation. Regarding the RAVDESS dataset [14], the following attributes were considered: Wav (audio signal), Modality (full-AV, video-only, audio-only), Vocal channel (speech, song), Emotion (neutral, calm, happy, sad, angry, fearful, disgust, surprised), Emotional intensity (normal, strong), Statement ("Kids are talking by the door," "Dogs are sitting by the door"), Repetition (1st repetition, 2nd repetition), and Actor (male, female). The RAVDESS dataset comprises a total of 1400 stimuli. To ensure the quality and consistency of the collected speech data, a series of preprocessing techniques were applied. These techniques encompassed silence removal and normalization, aiming to improve the dataset's overall usability and reliability.Similarly, for the TESS dataset [15], the following details were taken into account: A set of 200 target words were spoken in the carrier phrase "Say the word _____" by two actresses (aged 26 and 64 years). Recordings were made to represent seven emotions: anger, disgust, fear, happiness, pleasant (surprise), sadness, and neutral. Both actors were recruited from the Toronto area and possess English as their first language, university education, and musical training. Audiometric testing confirmed that their hearing thresholds fall within the normal range. The TESS dataset comprises a total of 2800 stimuli. By incorporating these databases into the research, a comprehensive and diverse set of emotional speech samples were utilized for training and evaluating models effectively. The applied preprocessing techniques aimed to enhance the dataset's quality and consistency, ensuring reliable and meaningful outcomes.

3.2 Exploratory Data Analysis (EDA)

EDA phase analyzes dataset to gain insights, identify patterns, and uncover relationships. Data is skewed due to lower "calm" samples but balanced using class weight and data augmentation techniques. **Figure 2** depicts the distribution of emotions in the data set.

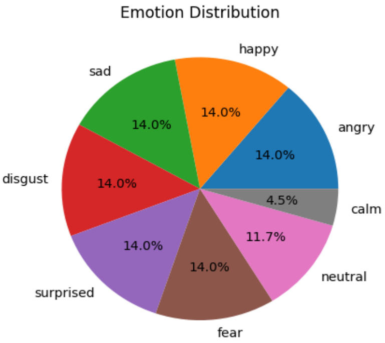

Figure 2. RAVDESS & TESS classes distribution

3.3 Data Augmentation

Data augmentation techniques improve model generalizability and address imbalanced emotion classes in the dataset data augmentation techniques such as:

3.3.1 **Noise**: Simulates real-world scenarios where there may be background noise as shown in **Figure 3**.

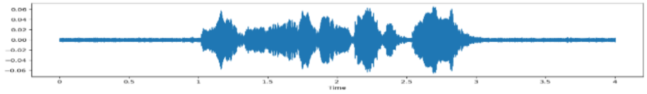

Figure 3. Noise

3.3.2 **Stretch**: The new sample has different durations than the original audio as shown in **Figure 4**.

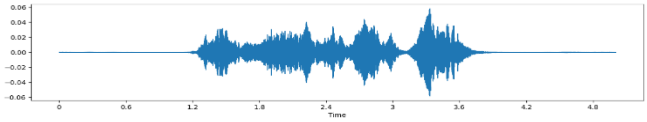

Figure 4. Stretch

3.3.3 **Shift**: The new sample has the same content as the original audio but is shifted in time as shown in **Figure 5**.

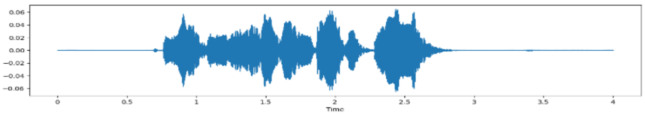

Figure 5. Shift

3.3.4 **Pitch**(frame/second): The new sample has a similar content to the original audio, but with a different pitch as shown in **Figure 6**.

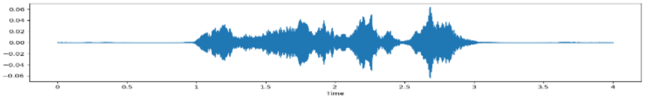

Figure 6. Pitch

3.3.5 **Higher speed**: To create new audio samples that have the same content as the original audio but are played back at a faster rate as shown in **Figure 7**.

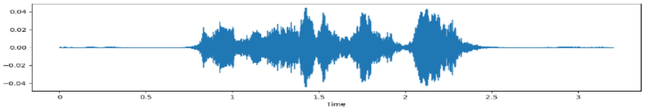

Figure 7. Higher speed

3.3.6 **Lower speed**: To create new audio samples that have the same content as the original audio but are played back at a slower rate as shown in **Figure 8**.

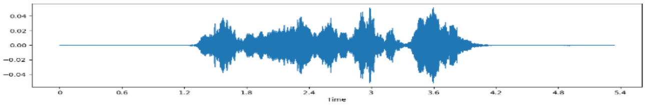

Figure 8. Lower speed

3.4 Features

Acoustic feature extraction plays a crucial role in SER by capturing the acoustic characteristics and patterns indicative of different emotional states [10]. The process uses LLDs and HLDs for feature extraction, capturing acoustic details and temporal variations in speech signals. LLDs use features like RMSE, chroma-STFT ,ZCR, MFCC, and Mel-Spectrogram, while HLDs summarize acoustic features' statistical properties and emotional characteristics. SER systems use LLDs and HLDs to classify and recognize different emotional states in speech.

3.4.1 Root Mean Square Energy (RMSE)

RMS Energy measures loudness by sample count and is less affected by outliers. **Figure 9** presents RMSE for neutral emotion. By taking the square root of the mean squared amplitude over a specific time interval, the RMSE is characterized as shown in Equation (1)

$$\mathrm{RMS}_t = \sqrt{\frac{1}{K}\sum_{k=t.K}^{(t+1)\cdot(K-1)} s\,(k)^2}. \quad (1)$$

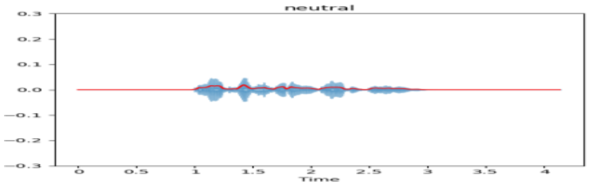

Figure 9. RMSE for neutral emotion

3.4.2 ZCR

ZCR represents the rate at which a signal crosses the zeroth line, indicating the frequency of positive/negative transitions. **Figure 10** presents ZCR for neutral emotion Mathematically, it can be expressed as shown in Equation (2) and (3).

$$\mathrm{ZCR} = \frac{1}{N-1}\sum_{n=1}^{N-1} \operatorname{sign}\left(s(n)\, s(n-1)\right), \tag{2}$$

where s = signal, N = length of a signal, and the sign(s(n) s(n-1)) is calculated as

$$\operatorname{sign}\left(s(n)\, s(n-1)\right) = \begin{cases} 1, \text{if } s(n)\, s(n-1) \geq 0 \\ 0, \text{if } s(n)\, s(n-1) < 0 \end{cases} \tag{3}$$

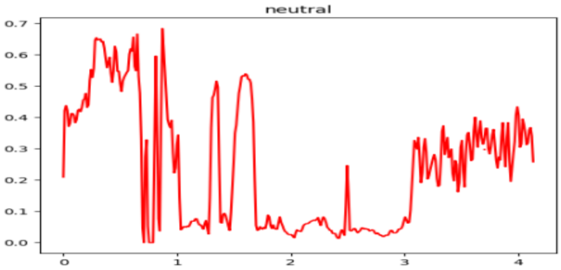

Figure 10. ZCR for neutral emotion

3.4.3 Mel-spectrogram

Mel-spectrogram displays frequencies in the Mel scale, a perceptually equal pitch spacing for human listeners, using the Fourier transform. **Figure 11** presents the creation of a Mel-spectrogram involving three primary steps: Compute the fast Fourier transform (FFT), Generate Mel scale and Generate spectrogram.

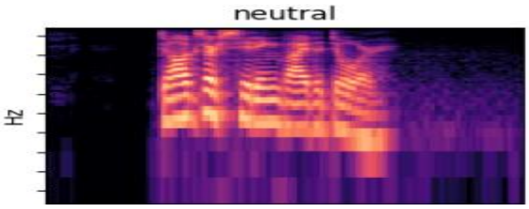

Figure 11. Mel-Spectrogram for neutral emotion

3.4.4 MFCC

The MFCC algorithm is a popular feature parameter in automatic speech and speaker recognition, capturing emotional characteristics in the frequency domain of the Mel scale. It is designed to address human hearing limitations, using linear spacing for frequencies below 1000 Hz and logarithmic spacing for frequencies above 1000 Hz. The algorithm also includes a subjective pitch element to capture essential phonetic characteristics in speech.[16]. **Figure 12** presents MFCC for neutral emotion

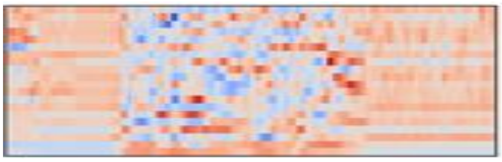

Figure 12. MFCC for neutral emotion

3.4.5 Chroma-STFT

The chroma-STFT algorithm calculates STFT and maps spectrum to 12-dimensional pitch class feature vector in Western musical scale. **Figure 13** presents Chroma-STFT for neutral emotion.

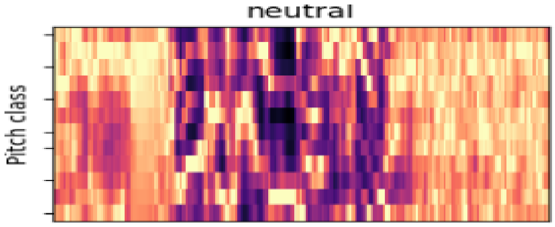

Figure 13. Chroma-STFT for neutral emotion

3.5 CUDA-DNN-LSTM (CU-DNN-LSTM) Architecture

The CU-DNN-LSTM DL model was implemented using the Keras library. The model is designed to address a classification task with multiple emotion classes. Here is a detailed description of the model:

3.5.1 **Input**: The model expects input data with a shape of (1, 386), representing a single sample of 386 features.

3.5.2 **Batch Normalization**: The input data is first normalized using Batch-Normalization to standardize the feature distribution.

3.5.3 **CU-DNN-LSTM Layers**: The model uses three stacked CU-DNN-LSTM layers for GPU acceleration, with the first having 512 units and returning sequences.

3.5.4 **Dropout Regularization**: Dropout is applied after each CU-DNN-LSTM layer with a rate of 0.5 to prevent overfitting by randomly dropping out units during training.

3.5.5 **Flatten Layer:** The output sequences from the CU-DNN-LSTM layers are flattened into a 2D representation.

3.5.6 **Dense Layers**: Flattened output undergoes connected Dense layer, Rectified Linear Unit (ReLU) activation function, dropout, and SoftMax for emotion classification probabilities, resulting in final classification probabilities.

3.5.7 **Regularization**: Both dense layers have kernel regularization with an L2 penalty (weight decay) of 0.01 to prevent overfitting.

3.5.8 **Compilation**: Model uses categorical-cross entropy loss function for multi-class classification, Adam optimizer, and accuracy metrics.

3.5.9 **Model Training**: The model trains using fit method on 50-epoch training data, with 32-batch size. Validation data monitors performance, early stopping based on accuracy, and class weights address data imbalance.

3.5.10 Hyperparameters like LSTM layer units, dropout rates, and regularization strengths were chosen using grid search to find the best configuration for a given task.

4. EXPERIMENTAL SETUP

Data is divided into three parts 10% validation,10% testing and 80% training.

4.1 Preprocessing: The trim function removes silence, data normalized using standard scaler, label encoded, and balanced using class weight method.

4.2 Features Extraction: The methodology divides audio files into smaller speeches, sections, frames, and LLDs, obtaining high-level feature vectors (HLDs) using statistical functions and DL models like recursive or convolutional networks. Librosa is a useful library for extracting features.

4.3 DL Models: Experiments in SER using the RAVDESS & TESS datasets, along with six data augmentation techniques, showed that the used DL models were effective and achieved an overall accuracy range of 47% to 91%. Extracting features using LLDs and HLDs, and comparing different models (MLP, CNN, GRU, LSTM, B-LSTM, and CUDNN-LSTM) revealed varying accuracy.

5. RESULTS

The SER study provides insights into the effectiveness and limitations of the proposed approach for emotion recognition from speech signals, addressing research objectives and discussing future directions. **Table 1** shows results of different DL models with above setups, results are test accuracy, F1 measure and recall.

Table 1 Accuracy results of different used DL Models

Model	Train	Valid	Test	F1-measure	Recall
CNN	86.77	86.56	81.71	0.82	0.82
GRU	96.37	88.92	88.21	0.88	0.89
LSTM	96.12	87.74	89.19	0.89	0.89
BLSTM	98.78	91.27	92.45	0.90	0.90
CU-DNN-LSTM	97.01	94.34	91	0.90	0.91

Among the models evaluated as shown at **Table 2**, the CU-DNN-LSTM stands out as the best performing model overall. While BLSTM achieved high trainability, validity, and test scores, indicating strong performance, it also showed signs of overfitting with a high train accuracy of 99% and a low validation accuracy 91% compared to CU-DNN-LSTM. On the other hand, the CU-DNN-LSTM demonstrated impressive generalization capabilities with high train scores (97.01%) and excellent valid scores (94.34%). It also achieved a competitive test score of 91%, indicating strong performance on unseen data. Therefore, the CU-DNN-LSTM model outperforms the other models, including the GRU, LSTM, and CNN, making it the preferred choice for the given task.

Table 2 F1-Score for each emotion

Emotion	Features		F1-Score
Neutral	**LLDs**:	**HLDs**:	0.90
calm	MFCC, Chroma, Mel-Spectrogram, Spectral contrast and Tonnetz	Mean and Standard deviation	0.84
happy			0.89
sad			0.94
angry			0.90
fear			0.89
disgust			0.90
Surprised			0.93

To represent how model fit the data while training Loss and Accuracy over training should be shown, as for **figure 14** it shows the loss over training and **figure 15** shows accuracy over training process.

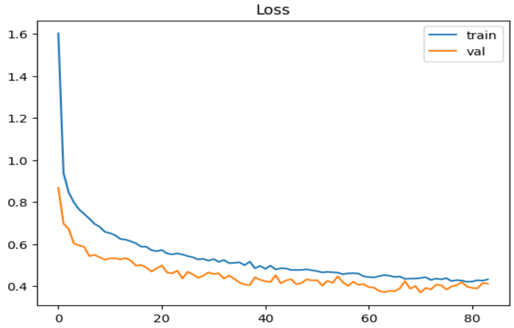

Figure 14. Loss in the training process

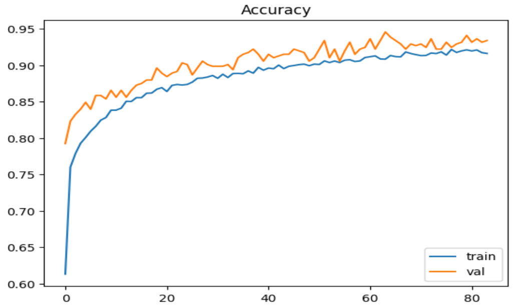

Figure 15. Accuracy in the training process

Figure 14 and **Figure 15** indicate that the model fits the data quite well and there is no high overfitting or underfitting indications. Comparing the used SER approach with existing methods provides insights into the strengths and weaknesses of the developed model.

Table 3 displays differences between the proposed system and previous systems in models, features, and test accuracy rates. The model excels in classifying sad and surprised emotions, while less accurate in classifying calm emotions. In the evaluation, the SER system outperformed previous methods. As it achieved a remarkable accuracy of 91%, outperforming most other systems in the last 4 years. This indicates the effectiveness of the proposed feature extraction techniques and classification algorithms in capturing and distinguishing emotional patterns in speech signals. However, it is important to note that direct comparisons between different studies may be challenging due to variations in dataset composition, feature extraction methods, and evaluation metrics. Thus, caution should be exercised when drawing direct conclusions from comparative analyses.

Table 3 SER performance with other systems

State of the art	Model	Features	Accuracy %
Hadhami Aouani et al [10].	SVM	MFCC, ZCR, TEO, HNR filtered by stacked AE	74.07
E Yu Shchetinin et al [17].	CNN+ BLSTM	Mel-cepstral, chroma, Spectral	74.80
Kannan Venkataramanan et al [18].	2D CNN with Global Avg. Pool	Log Mel Spectrogram	86.0
SER system	CUDNN-LSTM	MFCC, Chroma, Mel, contrast, Tonnetz	**91 %**

6. Conclusion

The study highlights limitations of the SER model, such as a single dataset, bias, and difficulty capturing nuanced emotions. A research project developed a SER system for detecting and classifying emotions in speech signals, achieving 91% accuracy. Future research should explore diverse datasets and integrate SER systems into real-world applications.

7. References

[1] A. Chiurco *et al.*, "Real-time Detection of Worker's Emotions for Advanced Human-Robot Interaction during Collaborative Tasks in Smart Factories," *Procedia Comput. Sci.*, vol. 200, pp. 1875–1884, Jan. 2022, doi: 10.1016/J.PROCS.2022.01.388.

[2] H. M. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for Speech Emotion Recognition," *Neural Networks*, vol. 92, pp. 60–68, Aug. 2017, doi: 10.1016/J.NEUNET.2017.02.013.

[3] E. Rejaibi, A. Komaty, F. Meriaudeau, S. Agrebi, and A. Othmani, "MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech," *Biomed. Signal Process. Control*, vol. 71, p. 103107, Jan. 2022, doi: 10.1016/J.BSPC.2021.103107.

[4] M. Rayhan Ahmed, S. Islam, A. K. M. Muzahidul Islam, and S. Shatabda, "An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition," *Expert Syst. Appl.*, vol. 218, p. 119633, May 2023, doi: 10.1016/J.ESWA.2023.119633.

[5] B. W. Schuller, "Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends," *Commun. ACM*, vol. 61, no. 5, pp. 90–99, May 2018, doi: 10.1145/3129340.

[6] "Emotion Recognition In Persian Speech Using Deep Neural Networks | Papers With Code." https://paperswithcode.com/paper/emotion-recognition-in-persian-speech-using (accessed Nov. 18, 2022).

[7] C. S. Kanani, K. S. Gill, S. Behera, A. Choubey, R. K. Gupta, and R. Misra, "Shallow over Deep Neural Networks: A Empirical Analysis for Human Emotion Classification Using Audio Data," pp. 134–146, 2021, doi: 10.1007/978-3-030-76736-5_13.

[8] "(PDF) Speech Emotion Recognition Using CNN." https://www.researchgate.net/publication/342231090_Speech_Emotion_Recognition_Using_CNN (accessed Nov. 18, 2022).

[9] X. Wu *et al.*, "Speech Emotion Recognition algorithm based on deep learning algorithm fusion of temporal and spatial features," *J. Phys. Conf. Ser.*, vol. 1861, no. 1, p. 012064, Mar. 2021, doi: 10.1088/1742-6596/1861/1/012064.

[10] H. Aouani and Y. Ben Ayed, "Speech Emotion Recognition with deep learning," *Procedia Comput. Sci.*, vol. 176, pp. 251–260, Jan. 2020, doi: 10.1016/J.PROCS.2020.08.027.

[11] X. Wu, W.-L. Zheng, and Z. Li, "Recognition of emotions in human speech with deep learning models You may also like Investigating EEG-based functional connectivity patterns for multimodal emotion recognition," *J. Phys. Conf. Ser. Pap. • OPEN ACCESS*, doi: 10.1088/1742-6596/1703/1/012036.

[12] M. M. Hussein *et al.*, "Human speech emotion recognition via feature selection and analyzing You may also like Image Pattern Recognition Algorithm Based on Improved Genetic Algorithm Qing Kuang-An Improved Artificial Neural Network Design for Face Recognition utilizing Harmony S," *J. Phys. Conf. Ser.*, vol. 1748, p. 42008, 2021, doi: 10.1088/1742-6596/1748/4/042008.

[13] Y. S. Lalitha, A. H. B. Sk, and M. V. A. Nag, "Neural Network Modelling of Speech Emotion Detection," *E3S Web Conf.*, vol. 309, p. 01139, 2021, doi: 10.1051/E3SCONF/202130901139.

[14] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)," Apr. 2018, doi: 10.5281/ZENODO.1188976.

[15] "Toronto emotional speech set (TESS) | TSpace Repository." https://tspace.library.utoronto.ca/handle/1807/24487 (accessed Jun. 25, 2023).

[16] A. Bala, A. Kumar, and N. Birla, "VOICE COMMAND RECOGNITION SYSTEM BASED ON MFCC AND DTW," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 12, pp. 7335–7342, 2010.

[17] E. Y. Shchetinin, "Recognition of emotions in human speech with deep learning models," *J. Phys. Conf. Ser.*, vol. 1703, no. 1, Dec. 2020, doi: 10.1088/1742-6596/1703/1/012036.

[18] "Emotion Recognition from Speech | Papers With Code." https://paperswithcode.com/paper/emotion-recognition-from-speech (accessed Nov. 18, 2022).