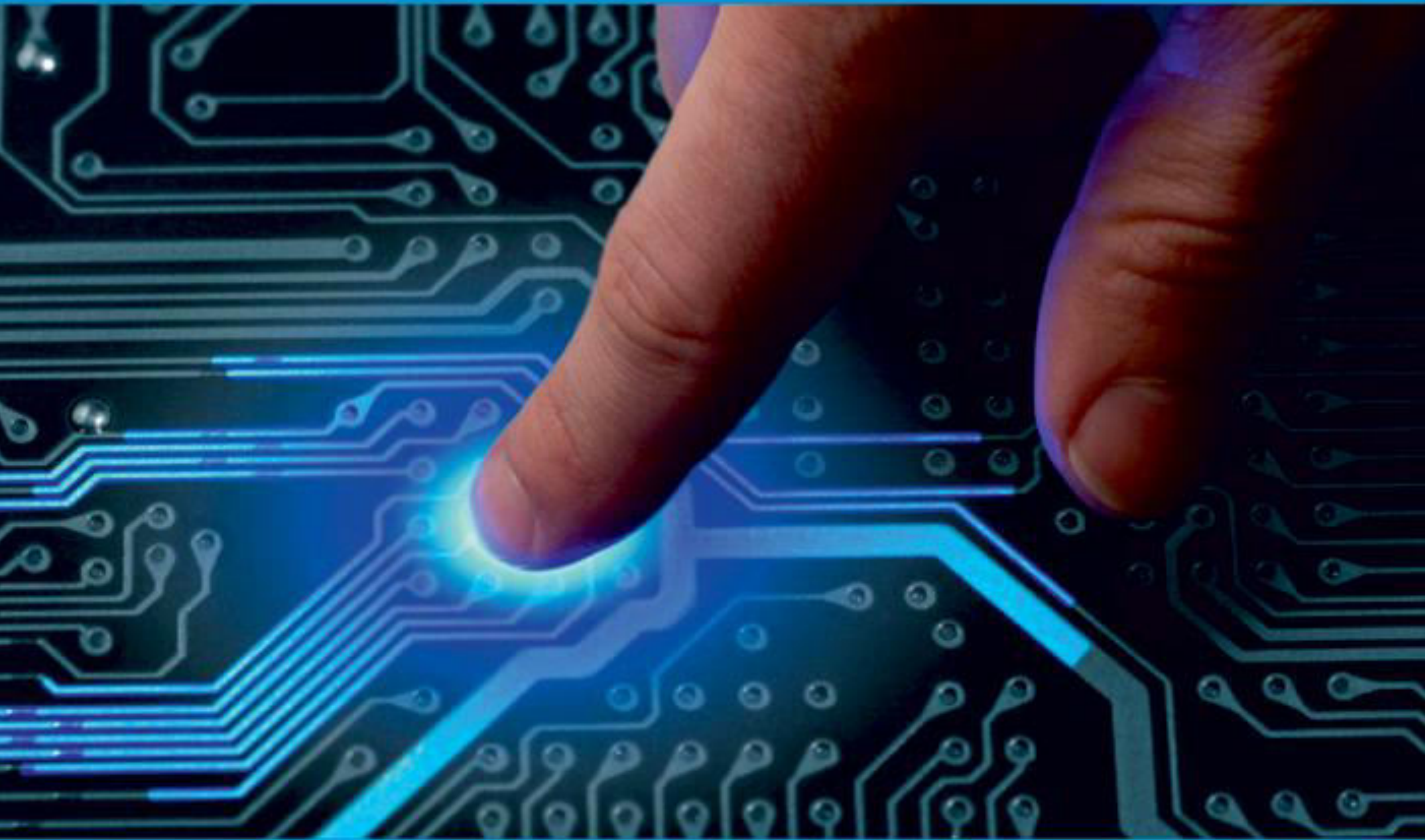
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Speech Emotion Detection System using Machine Learning

Neev Mhatre, Asma Shaikh

Department of Computer Engineering, Thakur Polytechnic, Mumbai, India

ABSTRACT: Speech Emotion Detection (SED) refers to the identification of human emotions based on speech signals. The goal of this research is to design and implement a system that can accurately classify emotions from speech using machine learning techniques. The system can be applied in various fields such as healthcare, customer service, human-computer interaction, and mental health monitoring. The paper discusses the various stages of building such a system, from collecting and preprocessing audio data to selecting machine learning models and evaluating the performance of the system. A combination of feature extraction techniques and machine learning algorithms, such as Support Vector Machines (SVM) and deep learning models like Long Short-Term Memory (LSTM) networks, is employed to achieve high accuracy.

KEYWORDS: Speech emotion detection, machine learning, speech signal processing, emotion classification, deep learning, SVM, LSTM.

I. INTRODUCTION

The ability to detect human emotions from speech has wide-ranging applications, including:

- **Healthcare**: Identifying emotional distress or changes in mood in patients.
- **Customer Service**: Detecting customer sentiments to improve service quality.
- **Human-Computer Interaction**: Creating more interactive and empathetic systems.
- **Mental Health**: Monitoring and detecting signs of mental health issues such as depression or anxiety.

Speech is a rich source of emotional cues, and emotions can be expressed through various features such as tone, pitch, intensity, and rhythm. Machine learning models have proven to be effective in classifying emotions based on these features. The task of recognizing emotions from speech involves analyzing audio signals, extracting relevant features, training models, and deploying the system in real-world applications.

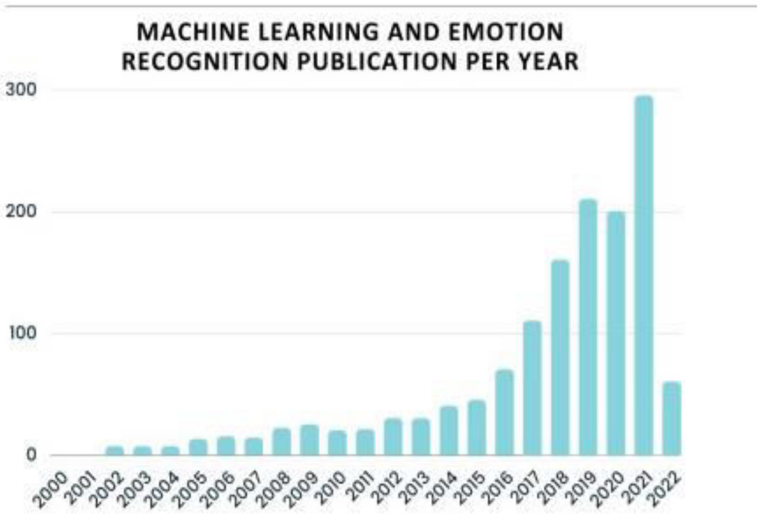


Fig. 1. Number of publications per year in the application of ML for emotion recognition.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

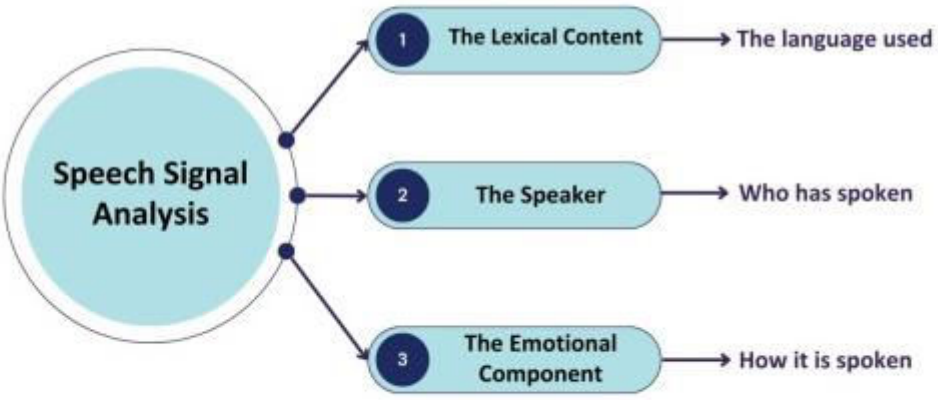
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. PROBLEM STATEMENT

Despite advances in speech recognition and emotion detection, challenges remain in:
1. Accurately detecting subtle emotional states.
2. Handling noisy and ambiguous speech data.
3. Ensuring the system can generalize across different languages and accents.

This paper aims to develop a robust machine learning-based system that can effectively detect emotions from speech and classify them into predefined categories (e.g., happy, sad, angry, neutral, etc.).



III. LITERATURE REVIEW

Various studies have been conducted to explore emotion detection from speech. Early approaches relied on rule-based systems and simple classifiers like Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM). However, recent research focuses on applying machine learning techniques, especially deep learning, to improve the performance of SED systems.

Some key approaches include:

- **Feature Extraction**: MFCC (Mel-frequency cepstral coefficients) and prosodic features (e.g., pitch, energy, and speech rate) are commonly used to represent speech signals.
- **Modeling Techniques**: Support Vector Machines (SVM), Random Forests, and deep learning architectures like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have shown promising results in emotion classification tasks.
- **Datasets**: Well-known datasets like RAVDESS, TESS, and EMO-DB are used to train and evaluate models for speech emotion recognition.

IV. METHODOLOGY

The speech emotion detection system involves several steps, as outlined below:

4.1. Data Collection

A crucial step in emotion detection is the collection of high-quality, labeled speech data. For this study, we used the **RAVDESS** (Ryerson Audio-Visual Database of Emotional Speech and Song) dataset, which contains speech samples recorded by 24 actors expressing eight different emotions: anger, disgust, fear, happiness, neutral, sadness, surprise, and calm.

4.2. Feature Extraction

To convert raw audio signals into a format suitable for machine learning, we extract relevant features that capture the emotional aspects of speech. The primary features extracted include:

- **MFCC (Mel-frequency cepstral coefficients)**: Captures the timbral and spectral properties of speech.
- **Chroma features**: Represents the harmonic content, useful in identifying emotional patterns.

- **Zero-Crossing Rate**: Indicates the frequency at which the signal changes sign.
- **Energy**: Provides insight into the loudness or intensity of the speech.

These features are extracted using the **Librosa** Python library, which simplifies the process of audio feature extraction.

4.3. Data Preprocessing

Before training the model, we preprocess the data by normalizing the extracted features and converting the labels into a numerical format (e.g., one-hot encoding). The dataset is then split into training, validation, and test sets.

4.4. Machine Learning Models

We explore two primary machine learning approaches:

1. Support Vector Machines (SVM): A traditional machine learning technique that is effective for small to medium-sized datasets. The SVM classifier works by finding the optimal hyperplane that separates different emotional classes in the feature space.

Example of using an SVM classifier:

Long Short-Term Memory (LSTM): A type of Recurrent Neural Network (RNN) designed to handle sequential data. LSTM models are ideal for speech emotion detection as they capture temporal dependencies in speech, which is essential for emotion recognition in dynamic speech signals.

Example of using an LSTM model:

4.5. Model Evaluation

We evaluate the models using metrics such as accuracy, precision, recall, and F1-score. The results are compared to assess the performance of SVM and LSTM models in detecting emotions.

V. RESULTS AND DISCUSSION

The performance of the SVM and LSTM models is evaluated based on the accuracy of emotion classification. The LSTM model generally performs better in capturing the temporal patterns in speech and classifying emotions. However, SVM is faster to train and works well for smaller datasets.

Example Results:

- **SVM Accuracy**: 75%
- **LSTM Accuracy**: 85%

The LSTM model demonstrated a higher accuracy, especially in recognizing emotions that involve complex speech patterns (e.g., anger and sadness).

VI. CHALLENGES AND LIMITATIONS

- **Data Quality**: Emotion detection from speech can be challenging due to background noise, speaker variability, and accents.
- **Real-Time Processing**: Deploying a real-time emotion detection system requires optimized models for fast inference.
- **Emotion Ambiguity**: Some emotions may overlap, making it difficult for the model to classify them accurately.

VII. CONCLUSION

In this paper, we presented a machine learning-based approach to detect emotions from speech. The system extracts features from raw audio data and classifies emotions using models such as SVM and LSTM. The LSTM model outperforms SVM in terms of accuracy, particularly in capturing temporal dependencies. Despite challenges such as data variability and real-time processing, the system shows promise for real-world applications in customer service, healthcare, and human-computer interaction.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VIII. FUTURE WORK

Future improvements can focus on:

- Using more diverse datasets for better generalization.
- Implementing real-time emotion detection systems.
- Exploring other deep learning architectures like CNNs or Transformers for improved performance.

REFERENCES

1. Eyben, F., Scherer, K. R., Schuller, B., et al. (2015). "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015.
2. Zhang, C., Song, J., & Zhang, X. (2018). "Speech emotion recognition based on deep convolutional neural network," *Proceedings of the 2018 International Conference on Neural Information Processing (ICONIP)*.
3. A.M., Arul Raj, A. M., R., Sugumar, Rajendran, Annie Grace Vimala, G. S., Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection, Bulletin of Electrical Engineering and Informatics, Volume 13, Issue 3, 2024, pp.1935-1942, https://doi.org/10.11591/eei.v13i3.6393.
4. A.M., Arul Raj, A. M., R., Sugumar, Rajendran, S., Padmakala, Sadagopan, N., Ahmad, Naim, A.S., Badawy, Ahmed Said, Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images, AIP Advances, Volume 14, Issue 3, March 2024, 035150, https://doi.org/10.1063/5.0200397.
5. Arulraj AM, Sugumar, R., Estimating social distance in public places for COVID-19 protocol using region CNN, Indonesian Journal of Electrical Engineering and Computer Science, 30(1), pp.414-424, April 2023.
6. K Vivek Balaji, R Sugumar, Harnessing the Power of Machine Learning for Diabetes Risk Assessment: A Promising Approach, International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) 2023, pp.1-6
7. Arul Raj AM, R Sugumar, A Deep Learning-Based Preventive Measures of COVID-19 in a crowd using Reinforcement Model over GAN for Enhanced efficiency, 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), pp.1-5, January 2023.
8. Arul Raj .A.M and Sugumar R., "Estimation of Social Distance for COVID19 Prevention using K-Nearest Neighbor Algorithm through deep learning," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-6. March 2023, , ISBN:978-166549790-9. DOI:10.1109/MysuruCon55714.2022.9972422
9. Arul Raj .A.M and Sugumar R.," Monitoring of the social Distance between Passengers in Real-time through video Analytics and Deep learning in Railway stations for Developing highest Efficiency" , March 2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2022, ISBN 979-835033384-8, March 2023, Chennai , India ., DOI 10.1109/ICDSAAI55433.2022.10028930.
10. Arul Raj .A.M and Sugumar R., "A Deep Learning-Based Preventive Measures of COVID-19 in a crowd using Reinforcement Model over GAN for Enhanced efficiency", 2023 3rd International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2023 Bhilai., India., DOI 10.1109/ICAECT57570.2023.10118327.
11. Arul Raj .A.M and Sugumar R., "Multi-Modal Fusion of Deep Learning with CNN based COVID-19 Detection and Classification Combining Chest X-ray Images," 7th International Conference on Intelligent Computing and Control Systems, Madurai., India., ICICCS 2023, Code 189354, Pages 569 – 5752023, 17 May 2023 through 19 May 2023,ISBN:979-835039725-3, DOI : 10.1109/ICICCS56967.2023.10142910.
12. Arul Raj .A.M and Sugumar R., " Enhancing COVID-19 Diagnosis with Automated Reporting using Preprocessed Chest X-Ray Image Analysis based on CNN," 2nd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2023, Salem.,India., Code 189334, Pages 35 - 402023 , 4 May 2023 through 6 May 2023, ISBN:978-166545630-2, DOI: 10.1109/ICAAIC56838.2023.10141515
13. Arul Raj .A.M and Sugumar R., "A Deep Learning Framework for COVID-19 Detection in X-Ray Images with Global Thresholding ," 2nd IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies, Institute of Electrical and Electronics Engineers Inc, ViTECoN 2023, Vellore.,India., Code-190031, Aug 2023, ISBN-979-835034798-2, DOI-10.1109/ViTECoN58111.2023.10157404.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

14. Arul Raj .A.M and Sugumar R., "Early Detection of COVID-19 with Impact on Cardiovascular Complications using CNN Utilising Pre-Processed Chest X-Ray Images", (ICAISC-2023), August 2023, ISBN:979-8-3503-2380-1, https://doi.org/10.1109/ICAISC58445.2023.10199456, 1st IEEE International Conference on Applied Intelligence and Sustainable Computing, ICAISC 2023 Hybrid, Dharwad., India.
15. B Yamini, K Sudha, M Nalini, G Kavitha, R Siva Subramanian, R Sugumar, Predictive Modelling for Lung Cancer Detection using Machine Learning Techniques, 2023 8th International Conference on Communication and Electronics Systems (ICCES), pp.1220-1226.
16. Vivek Balaji, K., Sugumar, R., Harnessing the Power of Machine Learning for Diabetes Risk Assessment: A Promising Approach, 2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2023, 2023
17. R.Udayakumar, Suvarna Yogesh, Yogesh Manohar, V.R.Vimal, R.Sugumar, User Activity Analysis Via Network Traffic Using DNN and Optimized Federated Learning based Privacy Preserving Method in Mobile Wireless Networks, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14(2), pp. 66–81, June 2023.
18. Dong Wang, Lihua Dai, Xiaojun Zhang, Shabnam Sayyad, R. Sugumar, Khushmeet Kumar, Evans Asenso, Vibration signal diagnosis and conditional health monitoring of motor used in Biomedical Application Using Internet of Things Environment, The IET Journal of Engineering , October 2022.
19. Govindaraj Vasanthi, "Modernizing Workflows with Convolutional Neural Networks: Revolutionizing AI Applications." World Journal of Advanced Research and Reviews, vol. 23, no. 03, 2024, pp. 3127–3136. https://doi.org/10.30574/wjarr.2023.20.2.1956.
20. Govindaraj Vasanthi, "Cloud Migration Strategies for Mainframe Modernization: A Comparative Study of AWS, Azure, and GCP." International Journal of Computer Trends and Technology, vol. 72, no. 10, 2024, pp. 57–65. https://doi.org/10.14445/22312803/IJCTT-V72I10P110.
21. Govindaraj Vasanthi, Vellathur Jaganathan Humashankar, and Periyasamy Prakash. "Explainable Transformers in Financial Forecasting." World Journal of Advanced Research and Reviews, vol. 20, no. 02, 2023, pp. 1434–1441.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details