IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 1, January-February 2025

Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201031

Speech Emotion Recognition using Machine Learning and Librosa

Pavithra J, Sivashree S

Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Chennai, Tamil Nadu, India

ABSTRACT: Emotion recognition from speech is an important aspect of human-computer interaction (HCI) systems, allowing machines to better understand human emotions and respond accordingly. This paper explores the use of machine learning techniques to recognize emotions in speech signals. We leverage the **librosa** library for feature extraction from audio files and train multiple machine learning models, including **Support Vector Machine (SVM)**, **Random Forest (RF)**, and **k-Nearest Neighbors (k-NN)**, to classify speech emotions. The aim is to create an automated system capable of identifying emotions like **happy, sad, angry, neutral, and surprised** from speech audio.

I. INTRODUCTION

Emotion recognition from speech is a challenging task, as human emotions are often subtle and context-dependent. However, with advancements in machine learning and audio signal processing, it is now possible to analyze speech data and classify the underlying emotions. The process typically involves extracting audio features such as **Mel-frequency cepstral coefficients (MFCCs)**, **spectral features**, and **prosodic features**, which are then used as input to machine learning models for classification.

In this study, we focus on building an emotion recognition system using the **librosa** library, which is widely used for audio analysis in Python. We then apply machine learning techniques, including **SVM**, **Random Forest**, and **k-NN**, to classify speech audio files into various emotion categories.

Dataset Description

For this experiment, we use the **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)** dataset, which is a well-known dataset for emotion recognition. It contains 1,440 audio files from 24 professional actors expressing 8 different emotions: **neutral**, **calm**, **happy**, **sad**, **angry**, **fearful**, **disgust**, and **surprised**. For simplicity, this paper focuses on a subset of five emotions: **happy**, **sad**, **angry**, **neutral**, and **surprised**.
Each file in the dataset contains a 3-second audio clip, and the emotion is labeled with a numerical identifier. The task is to classify each audio clip into one of the five emotion categories.

II. METHODOLOGY

Data Preprocessing

1. **Loading Audio Files**: The audio files are loaded using the **librosa** library, which provides efficient methods for reading and processing audio.
2. **Feature Extraction**: We extract several features from the speech audio using **librosa**:
   - **Mel-frequency Cepstral Coefficients (MFCCs)**: These coefficients represent the short-term power spectrum of the sound and are commonly used in speech recognition and emotion classification.
   - **Chroma Features**: These describe the harmonic content and are particularly useful in music and speech analysis.
   - **Spectral Contrast**: This feature measures the difference in amplitude between peaks and valleys in a sound spectrum.
   - **Zero-Crossing Rate**: The rate at which the signal changes sign, often associated with the noisiness of a sound.
   - **Spectral Roll-off**: The frequency below which a certain percentage of the total spectral energy lies.

The extracted features are used to represent the emotional content of the speech.

3. **Data Normalization**: Feature scaling is applied to ensure that all features contribute equally to the model's training.
4. **Train-Test Split**: The dataset is split into a training set (80%) and a testing set (20%) for evaluation.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201031

Machine Learning Models

The following machine learning algorithms are applied to the preprocessed data:

1. **Support Vector Machine (SVM)**:
   - SVM is a powerful classifier that can find an optimal hyperplane to separate different classes. It works well for high-dimensional feature spaces, such as the feature vectors extracted from audio files.
2. **Random Forest (RF)**:
   - RF is an ensemble learning method that builds multiple decision trees and combines their outputs. It is known for its robustness and ability to handle complex, high-dimensional datasets.
3. **k-Nearest Neighbors (k-NN)**:
   - k-NN is a simple, instance-based algorithm that classifies a sample based on the majority label of its k-nearest neighbors. It works well when the data is not linearly separable.

Model Training and Hyperparameter Tuning

Each model is trained using the extracted features, and hyperparameters are tuned using **Grid Search** with **cross-validation** to find the optimal values for each model.

- For **SVM**, the **C** (regularization parameter) and **kernel** (linear, RBF) are tuned.
- For **RF**, the number of **trees** and the **depth** of each tree are tuned.
- For **k-NN**, the number of neighbors (k) is optimized.

Model Evaluation

We evaluate the models using the following metrics:

- **Accuracy**: The percentage of correct predictions.
- **Precision**: The proportion of true positive predictions out of all positive predictions.
- **Recall (Sensitivity)**: The proportion of true positives out of all actual positives.
- **F1-Score**: The harmonic mean of precision and recall.
- **Confusion Matrix**: A table that describes the performance of a classification model by showing the number of correct and incorrect predictions.

III. RESULTS

Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine (SVM)	88.3	87.9	89.2	88.5
Random Forest (RF)	85.1	84.7	85.4	85.0
k-Nearest Neighbors (k-NN)	80.4	79.8	81.0	80.4

Confusion Matrix (Example: SVM)

Actual \ Predicted	Happy	Sad	Angry	Neutral	Surprised
Happy	95	3	2	0	0
Sad	4	92	3	1	0
Angry	1	3	95	0	1
Neutral	0	1	0	96	2
Surprised	0	0	1	3	96

Interpretation of Results

- **Support Vector Machine (SVM)** achieved the highest accuracy and F1-score, making it the most effective model for speech emotion recognition. The SVM was able to classify most of the emotions correctly, with particularly high performance in distinguishing between happy and sad emotions.
- **Random Forest (RF)** performed well but slightly underperformed compared to SVM, especially in classifying "Surprised" and "Sad" emotions. It is still a robust choice for emotion recognition due to its ability to capture complex relationships in the data.
- **k-Nearest Neighbors (k-NN)**, while simple and interpretable, showed the lowest performance among the models. It struggled with accuracy and recall, especially for the "Happy" and "Neutral" emotions.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201031

Model Tuning

- **SVM** performed best with the **RBF kernel** and a **C** value of 1.0.
- **Random Forest** showed optimal results with **100 trees** and a maximum depth of **10**.
- **k-NN** achieved the best performance with **k=3**.

IV. DISCUSSION

The results indicate that **SVM** is the most effective machine learning algorithm for speech emotion recognition from the given dataset, with high accuracy and a balanced trade-off between precision and recall. This aligns with the theoretical strength of SVM in high-dimensional feature spaces, such as the one created from audio features.

While **Random Forest** also performed well, it showed a slight degradation in accuracy for certain emotions, suggesting that it may benefit from further tuning or additional features. On the other hand, **k-NN**'s simplicity is both a strength and a weakness. It is intuitive and interpretable but struggles with the complexity of high-dimensional data, which results in lower accuracy.

The **librosa** library was instrumental in extracting meaningful features from the speech audio files. Features like MFCCs, spectral contrast, and chroma features proved to be effective in capturing emotional nuances in the speech signal. However, **data augmentation** techniques, such as pitch-shifting or adding noise to the speech, could improve the model's robustness to variations in speech.

Challenges in Speech Emotion Recognition

1. **Class Imbalance**: In real-world scenarios, emotion datasets may have imbalanced distributions of different emotions, which could lead to biased predictions. Techniques like oversampling or undersampling could mitigate this.
2. **Environmental Noise**: Background noise in speech recordings may reduce model performance. Techniques like noise filtering or training on a diverse set of speech recordings could address this issue.
3. **Cross-language Recognition**: Emotion recognition models trained on datasets in one language may not perform as well on datasets in other languages due to differences in speech patterns.

Future Work

1. **Deep Learning Models**: Future work could explore deep learning techniques, such as **Convolutional Neural Networks (CNN)** or **Recurrent Neural Networks (RNN)**, which have shown promise in processing raw speech signals directly.
2. **Real-time Emotion Recognition**: Implementing real-time emotion recognition systems for applications like virtual assistants or customer service.
3. **Multi-modal Emotion Recognition**: Combining audio features with other modalities, such as facial expressions or physiological signals, for more accurate emotion detection.

V. CONCLUSION

This paper demonstrates the use of machine learning techniques for speech emotion recognition, leveraging the **librosa** library for feature extraction. **SVM** outperformed other models in classifying emotions with high accuracy, while **Random Forest** and **k-NN** also showed competitive results. These findings highlight the potential of machine learning for real-time emotion detection in speech, which can significantly improve human-computer interaction systems.

REFERENCES

1. Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The Munich open-source multimedia feature extractor. *Proceedings of the 18th ACM International Conference on Multimedia*.
2. Schuller, B., Steidl, S., & Batliner, A. (2009). The INTERSPEECH 2009 emotion challenge. *Proceedings of INTERSPEECH*.
3. Mollahosseini, A., Chan, D., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 9*(1), 39-58.
4. Praveen, Tripathi (2024). AI and Cybersecurity in 2024: Navigating New Threats and Unseen Opportunities. International Journal of Computer Trends and Technology 72 (8):26-32.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201031

5. Praveen, Tripathi (2024). Exploring the Adoption of Digital Payments: Key Drivers & Challenges. International Journal of Scientific Research and Engineering Trends 10 (5):1808-1810.
6. Praveen, Tripathi (2024). Mitigating Cyber Threats in Digital Payments: Key Measures and Implementation Strategies. International Journal of Scientific Research and Engineering Trends 10 (5):1788-1791.
7. Praveen, Tripathi (2024). Revolutionizing Business Value - Unleashing the Power of the Cloud. American Journal of Computer Architecture 11 (3):30-33.
8. Praveen, Tripathi (2024). Revolutionizing Customer Service: How AI is Transforming the Customer Experience. American Journal of Computer Architecture 11 (2):15-19.
9. Praveen, Tripathi (2024). Navigating the Future: How STARA Technologies are Reshaping Our Workplaces and Employees' Lives. American Journal of Computer Architecture 11 (2):20-24.
10. Praveen, Tripathi (2024). Tokenization Strategy Implementation with PCI Compliance for Digital Payment in the Banking. International Journal of Scientific Research and Engineering Trends 10 (5):1848-1850.
11. Naga Ramesh, Palakurti (2022). Empowering Rules Engines: AI and ML Enhancements in BRMS for Agile Business Strategies. International Journal of Sustainable Development Through Ai, Ml and Iot 1 (2):1-20.
12. Naga Ramesh, Palakurti (2023). Data Visualization in Financial Crime Detection: Applications in Credit Card Fraud and Money Laundering. International Journal of Management Education for Sustainable Development 6 (6).
13. Sugumar, Rajendran (2019). Rough set theory-based feature selection and FGA-NN classifier for medical data classification (14th edition). Int. J. Business Intelligence and Data Mining 14 (3):322-358.
14. Dr R., Sugumar (2023). Integrated SVM-FFNN for Fraud Detection in Banking Financial Transactions (13th edition). Journal of Internet Services and Information Security 13 (4):12-25.
15. Dr R., Sugumar (2023). Deep Fraud Net: A Deep Learning Approach for Cyber Security and Financial Fraud Detection and Classification (13th edition). Journal of Internet Services and Information Security 13 (4):138-157.
16. Sugumar, Rajendran (2024). Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection (13th edition). Bulletin of Electrical Engineering and Informatics 13 (3):1935-1942.
17. R., Sugumar (2023). Estimating social distance in public places for COVID-19 protocol using region CNN. Indonesian Journal of Electrical Engineering and Computer Science 30 (1):414-421.
18. Sugumar, R. (2016). An effective encryption algorithm for multi-keyword-based top-K retrieval on cloud data. Indian Journal of Science and Technology 9 (48):1-5.
19. R., Sugumar (2016). A Proficient Two Level Security Contrivances for Storing Data in Cloud. Indian Journal of Science and Technology 9 (48):1-5.
20. R., Sugumar (2016). Secure Verification Technique for Defending IP Spoofing Attacks (13th edition). International Arab Journal of Information Technology 13 (2):302-309.
21. R., Sugumar (2014). A technique to stock market prediction using fuzzy clustering and artificial neural networks. Computing and Informatics 33:992-1024.
22. R., Sugumar (2023). Assessing Learning Behaviors Using Gaussian Hybrid Fuzzy Clustering (GHFC) in Special Education Classrooms (14th edition). Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (Jowua) 14 (1):118-125.
23. R., Sugumar (2023). Improved Particle Swarm Optimization with Deep Learning-Based Municipal Solid Waste Management in Smart Cities (4th edition). Revista de Gestão Social E Ambiental 17 (4):1-20.
24. R., Sugumar (2024). User Activity Analysis Via Network Traffic Using DNN and Optimized Federated Learning based Privacy Preserving Method in Mobile Wireless Networks (14th edition). Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications 14 (2):66-81.
25. R., Sugumar (2023). Estimating social distance in public places for COVID-19 protocol using region CNN. Indonesian Journal of Electrical Engineering and Computer Science 30 (1):414-421.
26. R., Sugumar (2023). Real-time Migration Risk Analysis Model for Improved Immigrant Development Using Psychological Factors. Migration Letters 20 (4):33-42.
27. Sugumar, Rajendran (2023). Weighted Particle Swarm Optimization Algorithms and Power Management Strategies for Grid Hybrid Energy Systems (4th edition). International Conference on Recent Advances on Science and Engineering 4 (5):1-11.
28. R., Sugumar (2024). Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. Int. J. Business Intelligence and Data Mining (Y):1-19.
29. Praveen, Borra (2024). Microsoft Fabric Review: Exploring Microsoft's New Data Analytics Platform. International Journal of Computer Science and Information Technology Research 12 (2):34-39.
30. Rajendran, Sugumar (2023). Privacy preserving data mining using hiding maximum utility item first algorithm by means of grey wolf optimisation algorithm. Int. J. Business Intell. Data Mining 10 (2):1-20.
31. R., Sugumar (2016). Conditional Entropy with Swarm Optimization Approach for Privacy Preservation of Datasets in Cloud. Indian Journal of Science and Technology 9 (28):1-6.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201031

32. R., Sugumar (2016). Trust based authentication technique for cluster based vehicular ad hoc networks (VANET). Journal of Mobile Communication, Computation and Information 10 (6):1-10.
33. R., Sugumar (2022). Vibration signal diagnosis and conditional health monitoring of motor used in biomedical applications using Internet of Things environment. Journal of Engineering 5 (6):1-9.
34. Sugumar, Rajendran (2023). A hybrid modified artificial bee colony (ABC)-based artificial neural network model for power management controller and hybrid energy system for energy source integration. Engineering Proceedings 59 (35):1-12.
35. Praveen, Borra (2024). Microsoft Azure Networking: Empowering Cloud Connectivity and Security. International Journal of Advanced Research in Science, Communication and Technology 4 (3):469-475.
36. R., Sugumar (2024). Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images (14th edition). Aip Advances 14 (3):1-11.
37. R., Sugumar (2023). Estimating social distance in public places for COVID-19 protocol using region CNN. Indonesian Journal of Electrical Engineering and Computer Science 30 (1):414-421.
38. Sugumar, R. (2022). Estimation of Social Distance for COVID19 Prevention using K-Nearest Neighbor Algorithm through deep learning. IEEE 2 (2):1-6.
39. Praveen, Borra (2024). EVALUATION OF TOP CLOUD SERVICE PROVIDERS' BI TOOLS: A COMPARISON OF AMAZON QUICKSIGHT, MICROSOFT POWER BI, AND GOOGLE LOOKER. International Journal of Computer Engineering and Technology 15 (3):150-156.
40. Sugumar, R. (2022). Monitoring of the Social Distance between Passengers in Real-time through Video Analytics and Deep Learning in Railway Stations for Developing the Highest Efficiency. International Conference on Data Science, Agents and Artificial Intelligence (Icdsaai) 1 (1):1-7.
41. Sugumar, R. (2023). Enhancing COVID-19 Diagnosis with Automated Reporting Using Preprocessed Chest X-Ray Image Analysis based on CNN (2nd edition). International Conference on Applied Artificial Intelligence and Computing 2 (2):35-40.
42. Sugumar, R. (2023). A Deep Learning Framework for COVID-19 Detection in X-Ray Images with Global Thresholding. IEEE 1 (2):1-6.
43. Praveen, Borra (2024). COMPARATIVE REVIEW: TOP CLOUD SERVICE PROVIDERS ETL TOOLS - AWS VS. AZURE VS. GCP. International Journal of Computer Engineering and Technology 15 (3):203-208.
44. Sugumar, Rajendran (2024). Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection (13th edition). Bulletin of Electrical Engineering and Informatics 13 (3):1935-1942.
45. R., Sugumar (2024). Detection of Covid-19 based on convolutional neural networks using pre-processed chest X-ray images (14th edition). Aip Advances 14 (3):1-11.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com