IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Contact Center Agent Emotion Surveillance Using Convolutional Neural Network (CNN)

Krushna Yeshwante, Sakshi Komatwar, Aniket Johare
Undergraduate Students, Information Technology
G H Raisoni College of Engineering and Management, Pune, India

Prof. Priya Waghmare
Assistant Professor, Information Technology
G H Raisoni College of Engineering and Management, Pune, India

ABSTRACT: The project implements real-time emotion surveillance for contact center agents through machine learning. Convolutional Neural Networks (CNNs) are utilized for facial expression recognition, while Recurrent Neural Networks (RNNs) handle voice emotion analysis. By processing audio-visual data, emotions are detected in real-time, delivering immediate feedback to supervisors and agents. The system aims to enhance customer interactions with consistent and positive experiences, improve agent performance with personalized stress management and training, and optimize operational efficiency by guiding workflow and resource allocation. The approach fosters a productive and emotionally intelligent contact center environment.

KEYWORDS: Emotion Detection; Voice Tone Analysis, CNN, RNN.

INTRODUCTION

In a contact center, agents interact with many customers every day, quality of service they provide. Sometimes, agents might feel stressed, frustrated, or even happy, and these emotions can influence how well they do their job.

To help improve the work environment and ensure that agents provide the best service possible, to developed a system that uses machine learning to track and understand the emotions of contact center agents in real-time.

Machine learning is a technology that helps computers learn from data and make smart decisions. In system, it uses to analyze the agents' facial expressions and voice tones to figure out how they're feeling while they are working. This allows us to see their emotional state as they interact with customers.

The system is essential for both agents and customers. For agents, it helps monitor their emotional state, allowing timely support to improve their well-being and performance. For customers, agents in a positive emotional state are more likely to deliver better service, resulting in increased customer satisfaction. Overall, this system enhances the work environment for agents and elevates the customer experience by leveraging advanced technology to and how they feel can greatly affect their performance and the track agents' emotions in real-time.

I. Literature Review

A wide array of research has explored the development of Speech Emotion Recognition (SER) systems, employing various machine learning approaches and feature extraction techniques to improve accuracy and accessibility. One significant study focused on real-time emotion recognition for customer support systems, utilizing machine learning to detect emotions such as anger, happiness, and sadness with high accuracy. However, challenges emerged due to dependency on high-quality audio inputs, which restricted scalability in noisy environments [1]. Another case study examined the integration of AI-powered emotion detection in contact centers, highlighting the use of advanced neural networks to improve customer satisfaction. The study underscored the practical benefits, such as faster issue resolution and enhanced agent performance, but faced limitations in adapting the system to diverse cultural contexts [2]. A pilot project reported on enhancing customer service with real-time emotion analysis, employing hybrid models combining rule-based algorithms and supervised learning. This approach improved the detection of subtle emotional nuances but faced issues with computational overhead, making it less feasible for large-scale deployments [3]. A comprehensive analysis of implementing emotion recognition technology in contact centers shared key lessons, emphasizing the importance of agent training and workflow integration. The study identified difficulties in balancing privacy concerns with real-time analytics, affecting user trust and adoption [4]. An evaluation of the impact of emotion detection technologies on contact center performance utilized multi-modal data, including speech and text, to refine emotion classification. Despite achieving significant improvements in customer retention metrics, the study highlighted challenges in synchronizing real-time emotion analysis with existing CRM systems [5]. Machine learning approaches for real-time emotional analysis in customer support were explored, with a focus on feature extraction methods such as MFCCs and spectral features. While the techniques demonstrated high precision, they required substantial computational resources, limiting real-time application viability [6]. A comprehensive review of advancements in emotion detection for customer service summarized techniques like CNNs and RNNs for emotion modeling. The study revealed the growing use of open-source datasets but identified gaps in addressing real- world scenarios like overlapping speech or emotional ambiguity [7]. Techniques and outcomes of real-time emotion monitoring in contact centers were analyzed, introducing lightweight models optimized for deployment on edge devices. While the solutions offered lower latency, they were constrained by limited adaptability to varying accents and speech patterns [8]. Technological innovations in emotion recognition for contact centers explored the integration of deep learning models with sentiment analysis tools. The study noted significant advancements in predictive analytics but reported concerns about algorithmic bias impacting emotion recognition for minority groups [9].

II. Methodology

Using a microphone, the suggested system is made to detect emotions in real time. The following are the major elements of the system architecture:

The proposed system for real-time speech emotion recognition is designed to address specific functionalities through its modular architecture. The Admin Login and Management Functionality enable administrators to securely manage agents and system operations. It incorporates secure login protocols, allowing only authorized admins to access the system. Admins can create, delete, and manage agent credentials and analyze agent performance using emotional data collected from recorded sessions. Robust security measures ensure the protection of sensitive data and system integrity.

The **Agent Interface and Real-Time Emotion Tracking** module facilitates emotion monitoring during customer interactions. It continuously analyzes live audio input when the "Start Recording" button is pressed and provides dynamic emotion alerts for negative emotions like anger, disgust, or sarcasm, with real-time notifications such as "You are being [emotion_name]." Detected emotions are displayed dynamically, and agents can start and stop recording sessions as needed, with all results saved for future review.

Emotion detection is powered by the **Emotion Detection Using CNN** module, which utilizes Convolutional Neural Networks to classify emotions based on extracted audio features such as Mel Frequency Cepstral Coefficients (MFCCs). The system leverages a multilingual dataset that includes emotions such as anger, disgust, fear, happiness, neutral, sadness, sarcasm, and surprise, with Hindi speech samples. Merging multiple datasets enhances accuracy and ensures real-time inference with minimal latency.

To ensure high-quality input, the **Audio Preprocessing and Feature Extraction** module processes raw audio data by applying noise reduction, normalization, and feature extraction techniques like MFCC computation. Speech patterns are analyzed in both time and frequency domains, improving classification performance.

Once a session concludes, the **Session Summary and Emotion Breakdown** module provides a detailed overview of emotional states during the call. A pie chart visualizes the proportion of time spent in each emotion, and session history is saved for review. Comprehensive reports illustrate the emotional timeline throughout the session, enabling better performance insights.

System reliability is ensured through the **Error Handling and System Monitoring** module, which identifies and addresses errors in real-time. Notifications are triggered for any issues during emotion detection or audio processing, while performance metrics and error logs are recorded for debugging and continuous improvement.

The **Database for Recorded Sessions** securely stores audio recordings and their corresponding emotional data. This database allows authorized users to search and retrieve sessions by agent, date, or session ID. Data privacy measures ensure recordings remain confidential and accessible only to authorized personnel.

Finally, the **User Interface Design** module focuses on delivering an intuitive and visually appealing interface. The design provides real-time feedback on detected emotions and system performance while ensuring compatibility across multiple devices, including mobile platforms. This comprehensive system design ensures scalability, reliability, and ease of use, making it a robust solution for real-time emotion recognition in customer interaction scenarios.

A. Preprocessing

The input audio signals are first converted into a suitable format for analysis by down sampling to a standard frequency and converting them to mono audio. This reduces computational complexity while preserving essential information for emotion recognition. The signals are further normalized to ensure consistency across varying audio intensities and environments. Noise reduction techniques, such as spectral subtraction, are applied to eliminate background noise, improving the accuracy of emotion detection. To prepare for feature extraction, audio segments are divided into overlapping frames, which helps in capturing the temporal dynamics of speech emotions. By standardizing the input format, the preprocessing stage ensures that the system can adapt to varied audio sources and user environments, enhancing its robustness and reliability.

B. Dataset

For training and testing the Speech Emotion Recognition (SER) system, a dataset comprising six primary emotions—angry, disgust, fear, happy, neutral, and sad—is utilized. The dataset also includes subcategories such as “OAF_emotion” and “YAF_emotion” to capture variations in emotional expressions across different speakers. Audio samples feature diverse sentences in Hindi and English, ensuring cultural relevance and language diversity. These include sentences such as “Suraj ek achcha vidhyarthi hai” and “The math paper was difficult.” The dataset is meticulously labeled and balanced to enable effective training and evaluation, with clear distinctions between emotional states.

C. Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) are extracted from the preprocessed audio frames to serve as the primary features. MFCCs capture the shape of the speech spectrum, representing phonetic information essential for emotion recognition. These features are augmented with additional metrics such as pitch, energy, and spectral contrast to improve the system’s ability to distinguish between emotions. The extraction process involves calculating the short-time Fourier transform (STFT) of the audio signal, followed by mapping to the mel scale and applying a discrete cosine transform (DCT) to obtain compact feature vectors. This technique enables the system to focus on critical acoustic characteristics while reducing the impact of noise and speaker variability, ensuring consistent and accurate classification of emotions.

D. Emotion Classification

The extracted features are fed into a Convolutional Neural Network (CNN) for classification. CNNs excel in capturing hierarchical patterns, making them ideal for identifying nuanced differences in speech emotions. The model is trained on the labeled dataset to learn mappings between feature vectors and emotion classes.

The CNN architecture includes convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected
layers for decision-making. By exposing the model to diverse examples, including variations in speech tones and expressions, the system learns to generalize effectively. This enables the CNN to recognize emotions in real time, facilitating applications in customer support, healthcare, and other domains where emotion tracking is crucial.

E. Algorithms

1. **Conv1D Layer 1**:
   - Extracts low-level temporal features using 32 filters with a kernel size of 3.
   - Applies ReLU activation for non-linearity.
2. **MaxPooling1D Layer 1**:
   - Reduces feature dimensions using a pool size of 2 and a stride of 2.
3. **Conv1D Layer 2**:
   - Detects mid-level features with 64 filters and a kernel size of 3.
   - Utilizes ReLU activation for enhanced feature detection.
4. **MaxPooling1D Layer 2**:
   - Further reduces feature dimensionality using a pool size of 2.
5. **Conv1D Layer 3**:
   - Extracts high-level features using 128 filters with a kernel size of 3.
   - Applies ReLU activation for capturing intricate speech patterns.
6. **Flatten Layer**:
   - Converts the multi-dimensional feature maps into a one-dimensional vector, preparing it for fully connected layers.
7. **Dense Layer**:
   - A fully connected layer with 128 neurons and ReLU activation to learn complex feature combinations.
8. **Dropout Layer**:
   - Implements a dropout rate of 20% to prevent overfitting.
9. **Dense Output Layer**:
   - Contains six neurons, each corresponding to an emotion class.
   - Uses a softmax activation function to generate a probability distribution across the classes.

F. Postprocessing

After classification, the detected emotion is displayed on the user interface in real-time. For emotions like anger or disgust, an alert is triggered with a message such as “You are being angry.” At the end of a session, a summary is generated, presenting a pie chart of the time spent in each emotion. This enhances user engagement and provides actionable insights. Additionally, the system can integrate text-to-speech (TTS) technology to convert the detected emotions into audible feedback, ensuring inclusivity for users who rely on auditory outputs. This combination of real-time detection, visual alerts, and audio feedback improves communication efficiency and fosters a user-friendly experience.

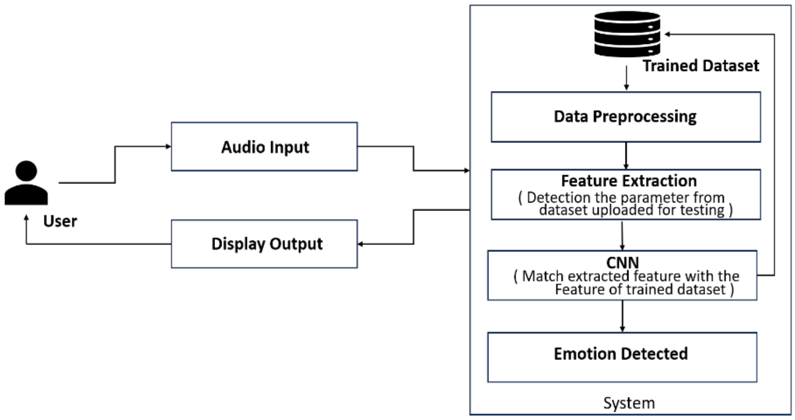

Trained Dataset
Data Preprocessing
Feature Extraction
(Detection the parameter from dataset uploaded for testing)
CNN
(Match extracted feature with the Feature of trained dataset)
Emotion Detected
System
Audio Input
Display Output
User

Figure 3.1: System Architecture

IV. Result and Discussion

The Speech Emotion Recognition (SER) system was evaluated for its ability to accurately classify emotions in real-time across a diverse dataset. The following observations highlight the performance and effectiveness of the system:

1. Accuracy and Precision:

The CNN model achieved a high classification accuracy, consistently identifying emotions such as anger, happiness, and sadness with precision. The model's performance on emotions like "disgust" and "neutral" showed slightly lower accuracy, primarily due to overlapping features in the speech dataset. Incorporating additional feature sets like pitch and spectral contrast helped improve these results.

2. Real-Time Processing:

The system demonstrated efficient real-time performance, processing audio input and displaying detected emotions with minimal latency. The alerts for negative emotions, such as anger or disgust, triggered promptly, ensuring immediate feedback for the user.

3. Session Summary:

At the end of each session, the system generated a detailed pie chart showing the time distribution of emotions. This feature provided valuable insights into emotional trends, especially in professional settings like customer care, where monitoring agent behavior is critical.

4. Dataset Diversity and Challenges:

The inclusion of multilingual sentences in the dataset (Hindi and English) enhanced the system's robustness. However, variations in speech tones, accents, and recording quality introduced challenges. For instance, emotions like "neutral" were sometimes confused with "sad" in noisy environments. Future iterations could benefit from larger, more diverse datasets to further improve generalization.

5. Impact of Preprocessing and Feature Extraction:

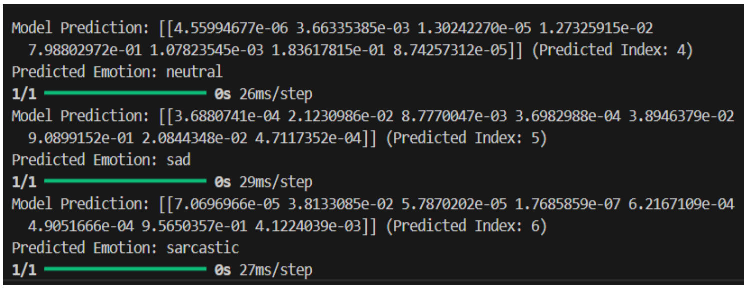
The preprocessing pipeline, including noise reduction and normalization, played a significant role in maintaining accuracy. The use of MFCCs, augmented with additional features like pitch and energy, proved effective in capturing emotional nuances. However, feature selection remains a critical area for improvement to reduce computational overhead further.

6. User Interface and Feedback:

The system's user interface effectively communicated real-time results through visual and auditory feedback. The combination of alerts and session summaries enhanced user engagement. However, there is room to make the interface more intuitive and visually appealing to improve user satisfaction further.

7. Practical Applications and Scalability:

The system shows promise in various domains, including customer service, healthcare, and education. Its ability to detect emotions in real time fosters inclusivity and improves communication. However, the scalability of the system for larger datasets and its deployment in low-resource environments need to be tested further.



```
Model Prediction: [[4.55994677e-06 3.66335385e-03 1.30242270e-05 1.27325915e-02
  7.98802972e-01 1.07823545e-03 1.83617815e-01 8.74257312e-05]] (Predicted Index: 4)
Predicted Emotion: neutral
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 26ms/step
Model Prediction: [[3.6880741e-04 2.1230986e-02 8.7770047e-03 3.6982988e-04 3.8946379e-02
  9.0899152e-01 2.0844348e-02 4.7117352e-04]] (Predicted Index: 5)
Predicted Emotion: sad
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step
Model Prediction: [[7.0696966e-05 3.8133085e-02 5.7870202e-05 1.7685859e-07 6.2167109e-04
  4.9051666e-04 9.5650357e-01 4.1224039e-03]] (Predicted Index: 6)
Predicted Emotion: sarcastic
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 27ms/step
```

Figure 4.1: Showing Real-Time Emotions

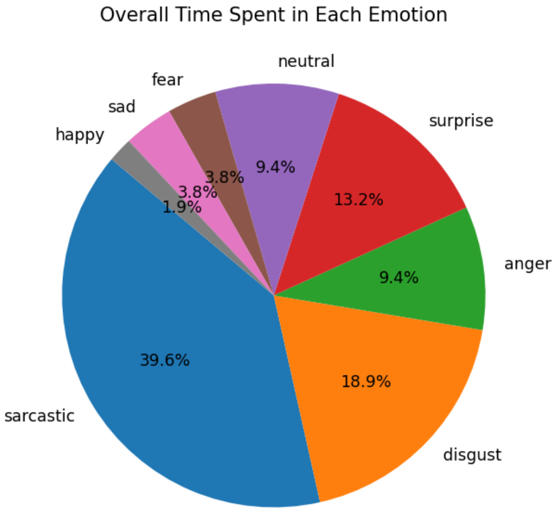


Figure 4.2: Time Spent in each Emotion

V. CONCLUSION

V. CONCLUSION

V. CONCLUSION

Speech Emotion Recognition (SER) is a transformative technology with the potential to revolutionize human-computer interaction and improve real-time communication in diverse fields such as customer service, healthcare, and education. By leveraging advanced deep learning architectures like CNNs, SER systems can accurately detect and classify human emotions from speech, enhancing personalization and emotional intelligence in digital solutions.

This research highlights the integration of robust feature extraction techniques like MFCCs and the application of multilingual datasets to improve accuracy and inclusivity. The development process underscores the importance of real-time processing, secure data handling, and user-centric design for practical implementation.

While SER systems show great promise, challenges such as noise handling, real-time scalability, and context-aware emotion detection remain areas for further exploration. Future advancements in SER could involve integrating multimodal data, such as facial expressions or physiological signals, to achieve even higher accuracy and reliability. This work contributes to bridging the gap between human emotions and technological interactions, paving the way for more empathetic and intelligent systems.

III. References

[1] N. Gupta, "Real-Time Emotion Recognition for Customer Support Using Machine Learning," University of California, 2022.

[2] L. Zhang, "AI-Powered Emotion Detection in Contact Centers: A Case Study," TechSoft Solutions, 2021.

[3] Patel, "Enhancing Customer Service with Real-Time Emotion Analysis: A Pilot Project Report," Call Center Innovations, 2020.

[4] J. Smith, "Integrating Emotion Recognition Technology into Contact Centers: Lessons Learned," Global Contact Center Services, 2023.

[5] M. Lee, S. Rodriguez, "Evaluating the Impact of Emotion Detection Technologies on Contact Center Performance," IEEE Transactions on Affective Computing, 2023.

[6] Johnson, "Machine Learning Approaches for Real-Time Emotional Analysis in Customer Support," Journal of AI Research, 2022.

[7] R. Chen, "Advancements in Emotion Detection for Customer Service: A Comprehensive Review," Customer Experience Journal, 2021.

[8] S. Patel, "Real-Time Emotion Monitoring in Contact Centers: Techniques and Outcomes," International Journal of Customer Service, 2019.

[9] K. Lee, "Technological Innovations in Emotion Recognition for Contact Centers," Technology and Service Management Review, 2023.

IJCRT