Grenze International Journal of Engineering and Technology, June Issue

Noise Removal, Feature Extraction and Classification of Environmental Sounds

Shobha K[1] and Rajashekhara S[2]
[1]Siddaganga Institute of Technology/Department of Computer Science & Engineering, Tumakuru, India
Email: shobhak@sit.ac.in
[2]Siddaganga Institute of Technology /Department of Chemical Engineering, Tumakuru, India
Email: rs@sit.ac.in

***Abstract*—The deep learning-based web application which aims to enhance the audio experience by removing unwanted audio and classifying environmental sounds will mainly focus on identifying specific environmental sounds and providing valuable insights and information about them through feature extraction techniques. Additionally, the webapp will offer real-time background noise reduction to improve the speaker's voice quality during online meetings. The web app will harness the capabilities of deep learning algorithms to analyze the audio data and utilizes advanced audio signal processing techniques to remove unwanted sounds. The application will have the ability to classify diverse environmental sounds, including traffic noise, bird chirping and water flowing. The users will have the option to upload their own recorded audio samples, which the web app will analyze and classify accordingly. By intelligently filtering out background noise, the web app will significantly enhance the quality of audio during online meetings, allowing participants to have clearer and more productive discussions.**

***Index Terms*— Classification of environmental sound, Noise removal, Noisy audio, Environmental data, Live noise reduction.**

I. INTRODUCTION

Sound classification holds an immense theoretical and practical significance. One of them being the development of a system that can identify specific environmental sounds and provide valuable insights and information about the acoustic surroundings. Sound classification is a widely researched area with many practical applications, such as speech recognition, audio indexing and audio content analysis. In an environment without a model to identify and categorize sounds, there are several sound classification problems that can arise. Firstly, distinguishing between different types of alarms becomes challenging, such as fire alarms, burglar alarms, or car alarms, which may require immediate attention for safety or security reasons. Secondly, identifying and categorizing animal sounds can be problematic, especially in situations where certain animal calls or vocalizations indicate danger or the presence of specific species. Thirdly, differentiating between various types of machinery noises is important for maintenance and troubleshooting purposes, as specific sounds may indicate malfunctions or performance issues [3].

The initial phase of developing a model to classify audio samples involves depicting the hierarchical structure of audio data. Secondly, three types of audio data features are to be extracted to construct the feature vector. Thirdly, discussion of how to classify audio data using the SVM classifier will be conducted.

Grenze ID: 01.GIJET.11.2.501

Environmental sounds encompass a wide range of auditory cues, including animal noises, traffic sounds, or natural phenomena. This classification capability will aid researchers, policymakers, and environmental enthusiasts in monitoring and understanding the acoustic characteristics of various environments. Removing unwanted sound from uploaded files or during online meetings is necessary since it serves several purposes. Firstly, it enhances the overall audio quality, ensuring clear and intelligible communication. By reducing background noise, such as ambient disturbances or microphone interference, the primary audio source becomes more prominent. Secondly, it helps to maintain professionalism and credibility, especially in business settings or professional recordings.

Eliminating unwanted sound signals a commitment to deliver high-quality content and creates a polished impression on the recipients. Lastly, noise reduction contributes to a more immersive and enjoyable user experience by eliminating distractions and creating a focused audio environment. Whether for personal or professional reasons, removing unwanted sound from files or online meetings significantly enhances the audio output, resulting in better communication and increased user satisfaction [5].

Online meetings are a form of virtual communication that allows people to connect and collaborate from different locations using internet-enabled devices. Online meetings can take various forms, including video conferencing, audio conferencing, webinars, and screen sharing sessions. They can be conducted in real-time or recorded for later viewing, depending on the needs of the participants. Online meetings have become increasingly popular due to their convenience, cost-effectiveness, and ability to bring people together regardless of their geographical location. They encompass everything from corporate meetings and educational workshops to training initiatives and social get-togethers. With the rapid advancement of technology, online meetings have become more accessible, enabling people to work together seamlessly from anywhere in the world.

A. Background Study

The task of sound classification involves attributing a class label to a given sound signal by examining its properties and characteristics. The goal of sound classification is to automatically categorize sounds into predefined classes, such as speech, music, or noise. Sound classification can be approached through different methods. Additionally, feature extraction is another widely used technique where significant features are derived from the sound signals and employed as input for machine learning algorithms. By automatically categorizing sounds into predefined classes, sound classification can improve the efficiency and performance of various sound processing systems and lead to new insights and understanding of the sound signals [4].

Feature extraction for noise removal involves extracting meaningful characteristics from noisy signals to enhance their quality. The objective is to identify crucial and distinctive signal features that remain unaffected by the noise, enabling their utilization in eliminating unwanted interference. Another popular method for noise removal is wavelet transform, which involves transforming the signals into a wavelet representation and then removing coefficients that correspond to the noise. The application of noise removal finds utility across various domains, including image processing, audio processing, and communication systems. In audio processing, noise removal can be used to reduce background noise in a recording. Feature extraction for noise removal is an important step in many signal processing algorithms and is often used as a preprocessing step before applying noise removal techniques. By reducing the dimensionality of the signals and isolating it, the most important feature extraction for noise removal can improve the performance of noise removal algorithms and lead to better results. Noise removal is a process in signal processing where unwanted or interfering signals are reduced or eliminated to enhance the signal's quality. The objective of noise removal is to separate the original, desired signal from the noise or interference that has been added to it [8].

Removing noise during online meetings is important as noise can degrade the quality and clarity of signals, whether they are audio signals, image signals, or data signals. By removing noise, the desired signal can be restored to its original quality, leading to better perception, interpretation, and analysis of the signal. Noise can interfere with the transmission and reception of information. By removing noise, the signal-to-noise ratio is increased, making it easier to extract and understand the intended message. This is particularly crucial in applications such as telecommunication, wireless networks, and audio/video streaming [2].

Background noise is a pervasive and disruptive element in various environments, such as airports. For instance, imagine waiting for a flight at the airport when an essential business call from a high-profile customer unexpectedly comes through. The surroundings are filled with a cacophony of sounds: chattering voices, departing airplanes, and frequent flight announcements. In this situation, the ability to remove or minimize background noise becomes crucial to ensure a clear and undisturbed conversation. This is where Active Noise Cancellation (ANC) technology proves its value, as it actively suppresses unwanted noise, allowing the user to focus on the call without being overwhelmed by the surrounding soundscape [5].

Another example would be of Online meetings. As the meeting begins, the team lead notices that some participants have background noise coming from their surroundings. One team member is in a busy café, another is working from a shared office space, and a few others have children or pets in the background. The noise from these environments makes it difficult to hear and understand each other clearly. This instance shows that noise removal is important to interact with people during online meetings [7].

II. Related Work

A novel deep CNN specification that incorporates adverse knowledge ensemble, leads to varied outcomes in environmental sound classification. When it is combined with the data augmentation, this mixture outperforms the original CNN structure without any augmentation. The study also analyzes the impact of each kind of variation on sound class accuracy [9] [22][26]. Classification of urban sound problems using CNN proposes an alternative approach, emphasizing the impact of input spectrograms on sound categorization performance. Boxplots and confusion matrices were utilized to visualize different inputs. The findings indicate that the time resolution index significantly affects categorization performance, with optimal results typically observed for inputs with moderate time resolution [25]. Saumya et.al, work has captured the speaker's voice and apply filters to eliminate any captured noise, providing a viable solution to address the issues caused by audio disturbances. The implementation process begins with the input of a noisy audio file, which is processed through a deep learning (DL) model designed inorder to detect the specific type of noise within the input audio. For training this Neural Network, a dataset called "UrbanSound8k" was utilized, which consists of 8732 sound excerpts labeled with the 10 most common noise classes, amounting to a total size of 6GB. A CNN model was employed to detect and remove the unwanted noises from the audio. The main research objective here was to construct an efficient model that can seamlessly be integrated into various audio-related applications [5]. Nugraha and colleagues conducted a comprehensive investigation into deep feed-forward neural networks for multi-channel source separation. They incorporated a multi channel Gaussian model to merge the source spectrograms, effectively harnessing the spatial information from the microphone array. The team extensively examined various loss functions and the model hyper-parameters, and one noteworthy discovery was the comparable performance of the standard mean-square error loss function [14]. DeLiang Wang presents a novel approach to the active noise control (ANC) called deep ANC. Unlike traditional methods that rely on the adaptive signal processing using least mean square algorithm, Deep ANC can be trained to cancel out noise and noisy speech regardless of the nature of the reference signal. To address potential latency issues in ANC systems, they introduce a delay-compensated strategy. The experimental findings demonstrate the effectiveness of deep ANC in reducing wideband noise and its ability to perform well with untrained noises [20]. Andong Li et.al initial analysis was conducted to examine the shared characteristics of artificial noise in DNN based speech enhancement methods. It was discovered that residual noise exhibited nonstationary behavior & contained significant energy levels that frequently surpassed noise masking threshold, resulting in enhanced speech signals that were bothersome to the human listeners. Furthermore, the subjective listening tests indicated that proposed strategies garnered preference percentages exceeding 60% [21]. Bernd Porr mentioned in his work about how he employs a deep neural network to achieve real-time simultaneous learning and noise reduction, eliminating the need for the conventional sequential approach of training and filtering. A thorough analysis was conducted on data collected from 20 subjects to detect electrode failure or significant external interference [19]. Dai et.al, work results demonstrate how the convolutional neural network architecture with waveforms enhances the precision of urban sound classification and improves the efficiency of the model by reducing the number of the parameters required. The model implementation is done using Python with Keras (2018) [2]. Lei Yang et al. model's effectiveness in distinguishing classes, utilized the UrbanSound8K dataset. Experimental results demonstrated that the method which was proposed significantly enhanced accuracy of classification. However, the authors primarily focused on the model's generality and did not incorporate feature fusion or customize model parameters based on dataset characteristics [2]. Hao Jiang method elucidates the fundamental dataset required for extracting video structures. The authors introduce audio classification scheme, achieving an overall accuracy rate exceeding 96%. Future work aims to enhance the classification method to discern additional audio classes. Furthermore, the authors plan to develop an efficient scheme for enhancing the video structures [20].

III. Proposed Method

Data Collection and Data Preprocessing: Obtained a dataset consisting of paired recordings: one set with noisy audio samples and their corresponding clean versions. Ensured that the dataset covers various types of noise,

noise levels, and audio characteristics. Segment the noisy and clean audio samples into fixed-length frames or time windows. Applied the preprocessing techniques such as resampling, normalization, and possibly augmentation to create a more diverse dataset.
Model Architecture: Designed a CNN architecture suitable for noise removal. It should have layers capable of learning noise patterns and denoising audio signals effectively.
Model Training and Model Evaluation: Experiment with training strategies like using a combination of noisy and clean audio as input or training in a multi-task learning framework. Evaluated the trained noise removal model's performance on the validation set by calculating objective metrics like Signal- to-Noise Ratio (SNR) improvement. Performed subjective listening tests to assess the perceived audio quality and the effectiveness of the denoising process.

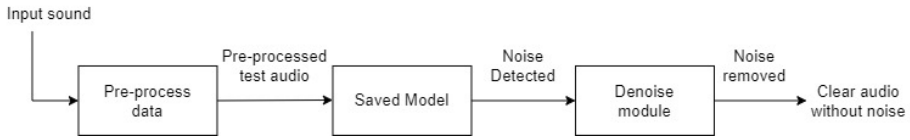

Fig.1 Noise Removal

Figure.1 depicts the detailed schematic representation of the activities performed in the process of noise removal. Collected data has been resampled and preprocessed and the audio will be saved. To remove the noise, a python denoise module has been used. Librosa which is python package and is a powerful audio processing library will provide various functionalities for audio analysis, including denoising capabilities. The application of this technology is prevalent in the domain of audio signal processing. After that clear audio without noise will be produced.
Figure.2, depicts the procedure involved for the process of extracting features. Upon receiving a sample sound from the dataset, it undergoes preprocessing through resampling, where the audio signal's sampling rate is adjusted. Following that, relevant features are extracted using MFCC and the resulting pattern is trained and stored in a database. In the recognition process, it will fetch the trained pattern from the database and recognize to give correct result.

Classification of Sound
Data Collection and Data Preprocessing: Obtained a diverse dataset of environmental sound recordings UrbanSound8K and ESC-50, including various classes such as traffic, birdsong, footsteps, etc. Ensured that the dataset covers a diverse range of audio characteristics, noise levels & environmental conditions to train a robust sound classification model. Preprocessed the audio data by segmenting it into fixed-length frames or time windows. Applied signal processing techniques like resampling, normalization, and possibly augmentation to enhance the dataset's variability.
Model Architecture: In this phase a CNN architecture suitable for sound classification is designed. Typically, it includes convolutional layers, activation functions, pooling layers, and fully connected layers. Compared the CNN model with RNN and LSTM model where the CNN model got more accuracy with less loss compared to other models.
Model Training and Model Evaluation: Monitor training progress, track metrics such as loss and accuracy, and apply regularization techniques to prevent overfitting. Evaluate the trained model's performance on the validation set by calculating metrics like accuracy and loss.
Fig.3, represents a resampling technique that can be used for sound classification to adjust the sample rate of a sound signal to a different value. Resampling aims to alter the sample rate of the audio signal, potentially impacting the frequency resolution and precision of the sound. To achieve this, the oversampling technique is employed. It involves an increasing number of instances in minority classes. The model performs Feature extraction to extract the relevant characteristics that improve the sound classification process. Mel-frequency cepstral co-efficient (MFCCs) are a popular feature extraction technique for sound classification. The resultant MFCCs effectively portray the timbral and spectral characteristics of the audio signal, frequently employed as input features in classification algorithms.
Noise filtering is a vital process, which finds application in sound classification by diminishing or eradicating undesirable background noise present in the sound signal. The primary aim of noise filtering is to enhance the sound signal's quality and facilitate precise identification of sound classes by sound classification algorithms. Low-pass filtering technique is used to attenuate high-frequency components in the audio signal while preserving the lower frequency components. Feature selection is a crucial step in sound classification, as it

determines which features of the sound signals will be used to train the classifier and make predictions. The objective of feature selection is to ascertain the most pertinent and enlightening characteristics of sound signals capable of effectively distinguishing between various sound categories. Utilizing a Convolutional Neural Network with 7 convolutional layers enables efficient way of feature selection. This process holds significance in enhancing the classifier's performance and accuracy. Environmental sound classification: After all these steps mentioned above, the sound will be classified and produced as a result.

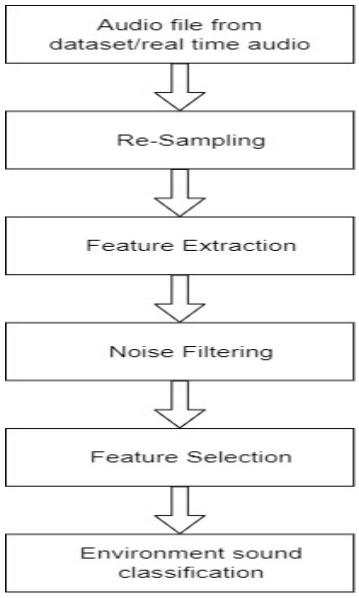

Fig 3. Block diagram of the sound classification

TABLE I. COMPARISON BETWEEN CNN, RNN & LSTM MODELS

Technique	Dataset	Advantage	Disadvantage
CNN	Urban Sound8k, ESC-50	It is capable of fetching higher-level features that are immutable to local spectral and time-related variations	Addresses the former shortcoming by learning filters that are displaced in both time and frequency, lacking however longer time-related context information
RNN	Urban Sound8k, ESC-50	Efficient in learning the long-term related context in the auditory signals	Does not easily capture the unvarying frequency in the domain, providing high-level modeling of the data and making it more difficult.
LSTM	Urban Sound8k, ESC-50	Less Data Loss than CNN, better performance and accuracy, a more generalized approach, and proves to be a reliable classifier for the networks that take magnitude Mel spectrograms	LSTM is more computationally intensive and prone to overfitting, although has less trainable parameters than CNN.

Real-time Audio Capture and Preprocessing: Utilize appropriate libraries or APIs to capture audio data in real-time from the microphone or audio stream during online meetings. Implement mechanisms to continuously receive audio frames or chunks as they are captured. Apply preprocessing steps to the incoming audio frames, such as resampling, normalization, or framing, to prepare them for further processing.

Noise Reduction and Adaptive Noise Reduction: Used the trained noise removal model to reduce background noise from the incoming audio frames. Feed the frames through the model, obtaining the denoised frames as the output. Applied the techniques like overlapping frames or time-domain filtering to ensure smooth and seamless noise reduction. Implemented the mechanisms to adaptively reduce noise based on the current noise level and

contextual information during the online meeting. Employed techniques such as noise estimation algorithms to adjust the noise reduction parameters in real-time.
Audio Output: Pass the denoised audio frames to the audio output device for transmission during the online meeting. Ensure proper synchronization with the video stream and other participants' audio.

IV. Experimental Results

UrbanSound 8k: It consists of 8,732 audio clips collected from various urban environments. These audio clips are categorized into 10 different classes, representing different urban sound types such as car horn, siren, street music, drilling, etc. Every audio clip included in the UrbanSound8K dataset conforms to the WAV format. The length of audio clips vary, typically spanning from 1 to 4 seconds.
ESC50: It consists of 2,000 environmental sound recordings, with each recording lasting approximately five seconds. The sounds cover a wide range of classes, including animals, natural sounds, human sounds, interior sounds, exterior sounds, and miscellaneous sounds. The dataset contains 50 sound classes, each representing a specific type of environmental sound. The classes include sounds like dog barking, rain, sea waves, door knock, helicopter, and more. The classes are evenly distributed, with 40 examples per class.
The application “NoizAway” incorporates advanced capabilities to classify and analyze various environmental sounds, providing valuable insights and information to the user. By leveraging real-time audio processing, the application intelligently filters out unwanted sounds such as chatter, machinery noise, or other disturbances, enabling the speaker's voice to stand out prominently.
Furthermore, the web application NoizAway goes beyond noise reduction by integrating a comprehensive sound classification system. Using a vast dataset of environmental sounds, the application employs sophisticated models to accurately recognize and categorize different types of sounds, ranging from natural elements like rain, wind, and birdsong, to urban sounds such as traffic, sirens, and construction noise. The classification functionality not only enhances user experience but also provides valuable insights into the surrounding environment. Users can access detailed information about the recognized sounds, including their origins, characteristics, and associated data. This feature is particularly useful for researchers, nature enthusiasts, and anyone interested in understanding and studying the auditory landscape. The application NoizAway offers customizable settings, allowing users to personalize noise reduction preferences and adjust sound classification parameters based on their specific needs and preferences.
The NoizAway webapp serves as a powerful tool for enhancing online communication by reducing background noise, improving voice clarity, and providing valuable information about various environmental sounds. With its advanced algorithms and intuitive interface, the application offers a seamless and immersive experience for users, revolutionizing the way virtual meetings are engaged and gaining insights from the auditory surroundings.

Table II. Actual and Predicted Outcomes of the Audio Sample

Actual	Prediction	Possibility Percentage
Dog_bark	Dog_bark	87.89822506904602
Car_horn	Car_horn	74.12785643209837
Air_conditioner	Air_conditioner	81.56178289235488
drilling	drilling	69.01782920197383

The above table represents the table of a few examples of actual sound present in the audio file and the corresponding audio that has been identified by the webapp along with their possibility percentage.
In the Figure 4.a and 4.b graphs represents the subplot function from the matplotlib.pyplot module to create a figure with four subplots arranged in a 2x2 grid. The figsize argument sets the size of the figure to 20 inches wide and 10 inches tall. The code then plots two signals on each subplot: a noisy signal and a clean signal. In Figure 4.a, the first subplot in the first row contains a plot of the 10th signal from the noisy_train dataset in red and the 10th signal from the clean_train dataset in blue in the 2nd row. In Figure 4.b the first subplot in the first row contains a plot of the 100th signal from the noisy_train dataset in red and the 100th signal from the clean_train dataset in blue in the 2nd row.
Fig 5 represents a graph indicating the accuracy values on y-axis and the epochs on x-axis. The lines indicate accuracy values for each model: CNN (74.09%), RNN (62.79%), and LSTM (60.42%).
The graph shows that CNN achieved highest accuracy of 74.09%, followed by RNN with an accuracy of 62.79%, and LSTM exhibiting the least accuracy of 60.42%.

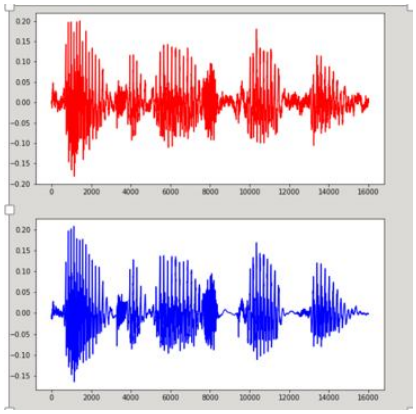

Fig.4.a 10th signal of Noisy and Clean Audio

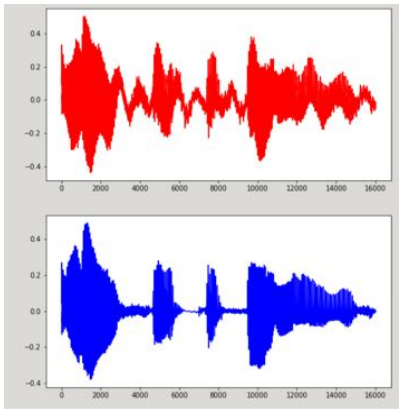

Fig.4.b 100 th signal of Noisy and clean audio

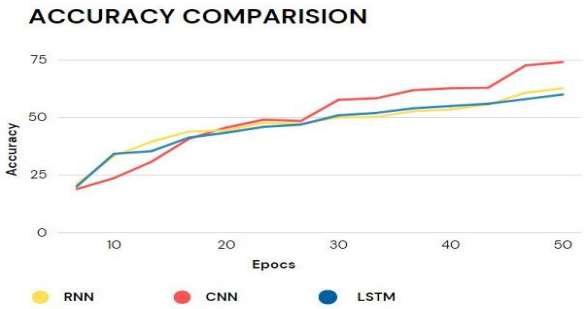

Fig 5 Accuracy comparison for Algorithms like CNN, RNN and LSTM

This comparison visually represents the performance differences among the models based on their accuracy, indicating that CNN outperformed both RNN and LSTM in this scenario.

Figure 6.a, 6.b & 6.c represents the graph of loss vs epoch obtained by using CNN, RNN & LSTM respectively. The number of epochs is depicted on the X-axis while loss is depicted on the Y-axis. Loss refers to value that represents how well the model is performing on the training data, while the validation loss (val_loss) indicates the model's performance on a separate validation dataset. The downward trend of the lines indicates that loss is decreasing as the number of epochs increases. This is a positive sign, as it suggests that the model is improving its performance over time and learning to make more accurate predictions.

Figure 7.a, 7.b & 7.c represents the graph of accuracy vs epoch obtained by using CNN, RNN & LSTM respectively. The number of epochs is depicted on x-axis while accuracy is depicted on y-axis. Accuracy indicates how well the model is performing on training data, while validation accuracy (val_accuracy) represents the model's performance on a separate validation dataset. The upward trend of the lines indicates that both the accuracy and val_accuracy will increase as number of epochs increases. This is a positive sign, as it suggests that the model is improving its performance over time and learning to make more accurate predictions.

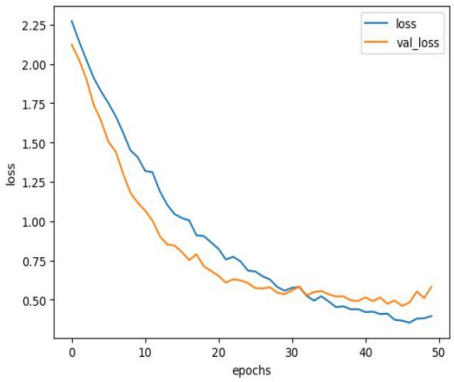

Fig 6.a Loss vs epoch using CNN

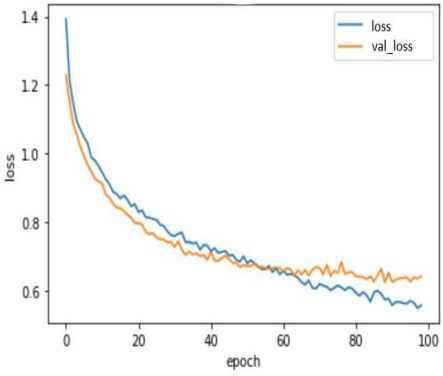

Fig 6.b Loss vs epoch using RNN

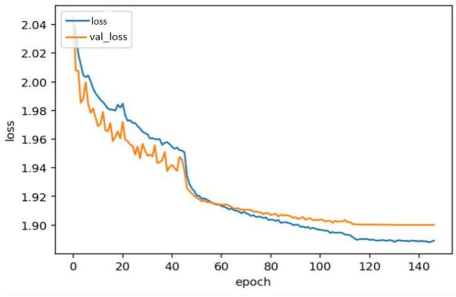

Fig 6.c Loss vs epoch using LSTM

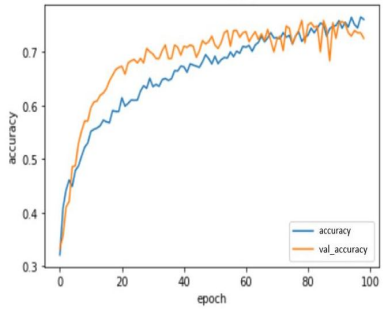

Fig 7.a Accuracy vs epoch using CNN

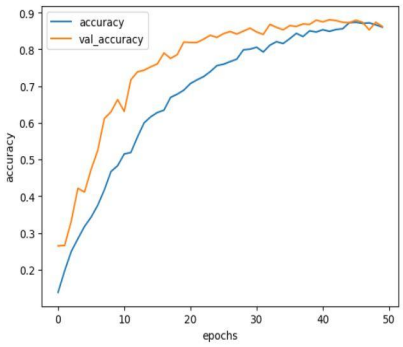

Fig 7.b Accuracy vs epoch using RNN

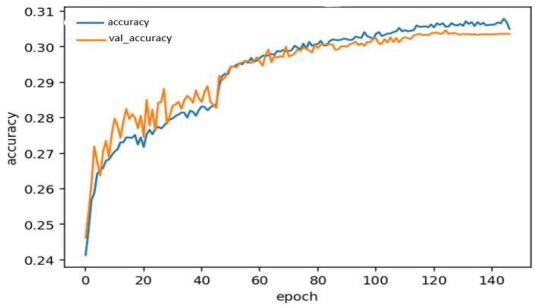

Fig 7.c Accuracy vs epoch using LSTM

V. Conclusion and Future Scope

The web application NoizAway enhances audio experiences by eliminating unwanted sounds and classifying environmental noises. It recognizes specific sounds, extracts key features, and improves speech clarity during online meetings through real-time noise reduction. Using deep learning algorithms, users can upload audio samples for classification, with advanced signal processing techniques filtering out noise. The app ultimately enhances communication and sound analysis across various applications.

A comparison of CNN, RNN, and LSTM models determined the most efficient algorithm for audio classification and noise removal. CNN achieved 74.73% accuracy, outperforming RNN (62.4%) and LSTM (60.42%), highlighting CNN's superior capability in improving audio quality and sound analysis.

The future of noise removal, feature extraction, and sound classification in web applications is promising. Advancements in machine learning and signal processing can enable real-time solutions for clearer audio and efficient sound analysis. Future improvements may refine algorithms to handle complex noise patterns and enhance deep learning-based denoising techniques.

For feature extraction, integrating methods like MFCCs and spectrograms with deep learning can capture finer audio details, improving classification accuracy. Optimizing computational efficiency through parallelization, hardware acceleration (GPUs), and model compression could enable seamless real-time noise reduction in virtual meetings, revolutionizing online communication.

References

[1] Baljinder Kaur & Jaskirat Singh, "Audio Classification: Environmental Sounds Classification," 2021.
[2] Yihua Yang, Ling Hong, Ke Liu & Chenwei Dai, "Urban Sound Source Classification and Comparison," 2022.
[3] Md. Adnan Habib, Zarif Raiyan Arefeen, Arafat Hussain, S. M. Rownak Shahriyer, Tanzid Islam & Rafeed Rahman, "Sound Classification Using Deep Learning for Hard of Hearing and Deaf People," 2022.
[4] Massoud Massoudi Software, Siddhant Verma & Riddhima Jain, "Urban Sound Classification Using CNN," 2021.
[5] Sainath Adapa, "Urban Sound Tagging Using Convolutional Neural Networks," 2019.
[6] Manmohan Dogra, Saumya Borwankar & Jayashree Domala, "Noise Removal from Audio Using CNN and Denoiser," 2021.
[7] Wenjie Muu, Bo Yin, Xianqing Huang, Jiali Xu & Zehua Du, "Environmental Sound Classification Using Temporal-Frequency Attention-Based Convolutional Neural Network," 2021.

[8] Srishti Garg, Tanishq Sehgal, Aakriti Jain, Yash Garg, Preeti Nagrath & Rachna Jain, “Urban Sound Classification Using Convolutional Neural Network Model,” 2019.
[9] J. Salamon & J. P. Bello, “Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification,” IEEE, 2017.
[10] Jivitesh Sharma, Ole-Christoffer Granmo & Morten Goodwin, “Environment Sound Classification Using Multiple Features,” 2019.
[11] I. Lezhenin, N. Bogach & E. Pyshkin, “Urban Sound Classification Using Long Short-Term Memory Neural Network,” in 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 57-60, 2019.
[12] T. N. Sainath, O. Vinyals, A. Senior & H. Sak, “Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks,” in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4580-4584, 2015.
[13] B. Zhu, K. Xu, D. Wang, L. Zhang, B. Li & Y. Peng, “Environmental Sound Classification Based on Multi-Temporal Resolution Convolutional Neural Network Combining with Multi-Level Features,” in Pacific Rim Conference on Multimedia, pp. 528-537, 2018.
[14] A. A. Nugraha, A. Liutkus & E. Vincent, “Multichannel Audio Source Separation with Deep Neural Networks,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 9, pp. 1652–1664, 2016.
[15] L. Pfeifenberger, M. Zohrer & F. Pernkopf, “DNN-Based Speech Mask Estimation for Eigenvector Beamforming,” in Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on, IEEE SigPort, 2017.
[16] J. Heymann, L. Drude, C. Boeddeker, P. Hanebrink & R. Haeb-Umbach, “Beamnet: End-to-End Training of a Beamformer-Supported Multi-Channel ASR System,” in Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on, 2017.
[17] C. Boeddeker, P. Hanebrink, J. Heymann, D. Lukas & R. Haeb-Umbach, “Optimizing Neural-Network Supported Acoustic Beamforming by Algorithmic Differentiation,” in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on, 2017.
[18] X. Anguera, “Beamformit, the Fast and Robust Acoustic Beamformer,” 2016.
[19] Bernd Porr, Sama Daryanavard, Lucía Muñoz Bohollo, Henry Cowan & Ravinder Dahiya, “Real-Time Noise Cancellation with Deep Learning,” 2022.
[20] Hao Zhang & DeLiang Wang, “Deep ANC: A Deep Learning Approach to Active Noise Control,” 2021.
[21] Yuxuan Ke, Andong Li, Chengshi Zheng, Renhua Peng & Xiaodong Li, “Low Complexity Artificial Noise Suppression Methods for Deep Learning-Based Speech Enhancement Algorithms,” 2021.
[22] J. Salamon & J. P. Bello, “Feature Learning with Deep Scattering for Urban Sound Analysis,” in the 2015 23rd European Signal Processing Conference (EUSIPCO), IEEE, pp. 724–728, 2015.
[23] J. Sang, S. Park & J. Lee, “Convolutional Recurrent Neural Networks for Urban Sound Classification Using Raw Waveforms,” in the 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, pp. 2444–2448, 2018.
[24] Karol J. Piczak, “Environmental Sound Classification with Convolutional Neural Networks,” in the 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), 2015.
[25] H. Zhou, Y. Song & H. Shu, “Using Deep Convolutional Neural Network to Classify Urban Sounds,” in TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, pp. 3089-3092, 2017. doi:10.1109/TENCON.2017.8228392.
[26] J. Salamon & J. P. Bello, “Unsupervised Feature Learning for Urban Sound Classification,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, pp. 171–175, 2017.