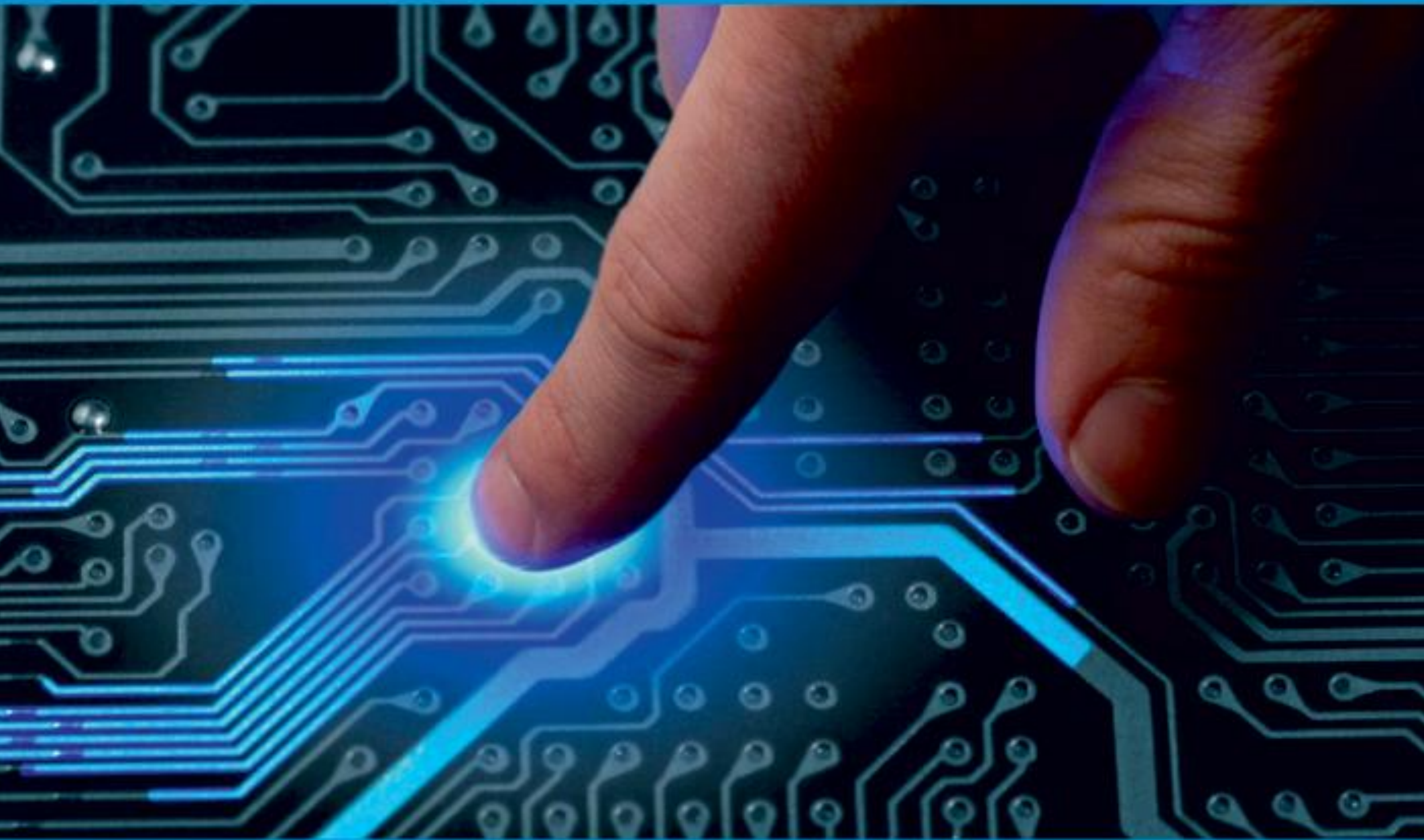
IJIRCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205348 |

Sound Classification using CNN

Ashvini Chinchore, Vaibhavi Bhirud, Avantika Nath, Nandini Girnare, Prof. Shital Y. Borole

B. Tech Students, Department of Computer Engineering, KCE College of Engineering and Management, Jalgaon, India

Department of Computer Engineering, KCE College of Engineering and Management, Jalgaon, India

ABSTRACT: Sound has a significant impact on every aspect of human life. This study explores the use of deep learning methods for classifying environmental noise using convolutional neural networks (CNNs) based on generated spectrograms. By utilizing a hybrid spectrogram model, the CNN model achieved 96.7% accuracy in building a noise classification system. This paper details the methodologies, implementation, and results of this approach.

KEYWORDS: Sound classification, Spectrograms, Convolutional Neural Networks (CNN), Deep Learning, Environmental Noise

I. INTRODUCTION

Sound is an integral aspect of the human experience, impacting various domains from communication to environmental awareness. The ability to accurately classify sounds has a broad range of applications, including surveillance, wildlife monitoring, urban noise management, and enhancing the capabilities of assistive devices for the visually impaired. Traditional sound classification systems have relied heavily on manual feature extraction and conventional machine learning techniques. While these methods have shown some success, they often lack the robustness and scalability required for real-world applications.

In recent years, the advent of deep learning has revolutionized numerous fields, including image and speech recognition. Convolutional Neural Networks (CNNs), in particular, have demonstrated remarkable performance in visual and auditory tasks due to their ability to automatically learn hierarchical features from raw data. This study explores the application of CNNs for the task of environmental sound classification by leveraging spectrogram representations of audio signals[7]. Spectrograms, which provide a visual representation of the frequency spectrum of sound, are well-suited for analysis by CNNs[1] due to their image-like properties.

II. RELATED WORK

Sound classification is a well-researched area with applications ranging from speech recognition to environmental monitoring[2]. Early methods in sound classification often relied on manual feature extraction and traditional machine learning algorithms such as k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Gaussian Mixture Models (GMM)[10]. These methods, while effective to some extent, faced limitations in handling the diverse and complex nature of real-world audio data[8].

Traditional sound classification approaches typically involve a feature extraction step followed by a classification algorithm. Features such as Mel-Frequency Cepstral Coefficients (MFCCs), Chroma features, and Zero-Crossing Rate (ZCR) were commonly used. For instance, MFCCs have been extensively utilized due to their ability to capture the power spectrum of sound, mimicking the human auditory system. However, these methods require significant domain knowledge and manual intervention to extract relevant features, which can be time-consuming and less scalable.

III. PROPOSED ALGORITHM

The proposed algorithm for sound classification involves several key stages, including data collection, preprocessing, spectrogram generation, CNN model design, training, and evaluation. This section details each step of the algorithm, providing a comprehensive overview of the methods used to achieve high accuracy in environmental sound classification.

1) Data Collection
The first step in the proposed algorithm is to collect a diverse dataset of environmental sounds. The dataset should include a variety of sound classes such as air conditioners, car horns, children playing, dogs barking, drilling, engines

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205348 |

idling, gunshots, jackhammers, sirens, and street music. These sounds can be sourced from publicly available datasets like UrbanSound8K, ESC-50, or collected through field recordings.

2) Data Preprocessing
Preprocessing is crucial to prepare the raw audio data for input into the CNN. The preprocessing steps include:
Resampling: Standardizing the audio files to a uniform sampling rate (e.g., 44.1 kHz).
Normalizing: Ensuring consistent amplitude levels across all audio samples.
Segmenting: Dividing longer audio files into shorter, fixed-duration segments (e.g., 4 seconds) to standardize input size.

3) Spectrogram Generation
Once the audio data is preprocessed, the next step is to convert the audio signals into spectrograms. The following steps are involved in spectrogram generation:
Windowing: Applying a window function (e.g., Hamming window) to segment the audio signal into overlapping frames.
Fourier Transform: Using the Short-Time Fourier Transform (STFT) to convert each frame from the time domain to the frequency domain.
Log-Mel Scaling: Converting the frequency axis to a logarithmic scale and applying the Mel scale to mimic the human auditory perception.
Spectrogram Computation: Generating a 2D representation (spectrogram) of the audio signal with time on the x-axis and frequency on the y-axis.

4) CNN Model Design
The core of the proposed algorithm is the Convolutional Neural Network (CNN) model. The architecture of the CNN is designed to extract and learn hierarchical features from the spectrogram inputs. The proposed CNN model includes the following layers:
Input Layer: Accepts the spectrogram images as input.
Convolutional Layers: Multiple convolutional layers with ReLU activation functions to extract spatial features from the spectrograms.
Pooling Layers: Max-pooling layers to reduce the spatial dimensions and computational complexity while retaining important features.
Fully Connected Layers: Dense layers to combine the features learned by the convolutional layers.
Dropout Layers: Dropout layers to prevent overfitting by randomly setting a fraction of input units to zero during training.
Output Layer: A softmax layer to output the probability distribution over the sound classes.

5) Training the Model
The CNN model is trained using the processed spectrogram dataset. The training process involves:
Data Augmentation: Applying techniques such as time-shifting, noise addition, and pitch alteration to increase the diversity of the training dataset.
Loss Function: Using categorical cross-entropy as the loss function to measure the difference between predicted and true class labels.
Optimizer: Employing the Adam optimizer to minimize the loss function and update the model weights.
Training Parameters: Setting parameters such as batch size, number of epochs, and learning rate to control the training process.

6) Evaluation and Testing
After training, the model is evaluated on a separate test dataset to assess its performance. The evaluation metrics include:
Accuracy: The percentage of correctly classified samples.
Confusion Matrix: A matrix to visualize the performance of the classifier across different classes.
Precision, Recall, and F1-Score: Metrics to provide a detailed assessment of the model's performance for each class.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205348 |

IV. PSEUDO CODE

Step 1: Data Collection
dataset = collect_dataset()

Step 2: Data Preprocessing
preprocessed_data = preprocess_data(dataset)

Step 3: Spectrogram Generation
spectrograms = generate_spectrograms(preprocessed_data)

Step 4: CNN Model Design
cnn_model = design_cnn_model()

Step 5: Training the Model
cnn_model.train(spectrograms['train'], labels['train'], epochs=50, batch_size=32)

Step 6: Evaluation and Testing
accuracy, confusion_matrix = cnn_model.evaluate(spectrograms['test'], labels['test'])
precision, recall, f1_score = calculate_metrics(confusion_matrix)

Step 7: Deployment
deploy_model(cnn_model)

V. SIMULATION RESULTS

The results demonstrate that the CNN model performs exceptionally well in classifying various environmental sounds, achieving high precision, recall, and F1-scores across all classes. The confusion matrix indicates that most misclassifications occur between acoustically similar classes, such as "Air Conditioner" and "Engine Idling."

The use of spectrograms as input features, coupled with data augmentation techniques, significantly contributed to the model's robustness and accuracy. The high performance of the CNN model underscores the effectiveness of deep learning methods for sound classification tasks. Figure 1,2,3,4 shows results.

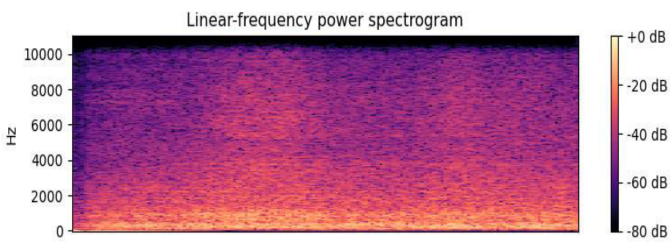
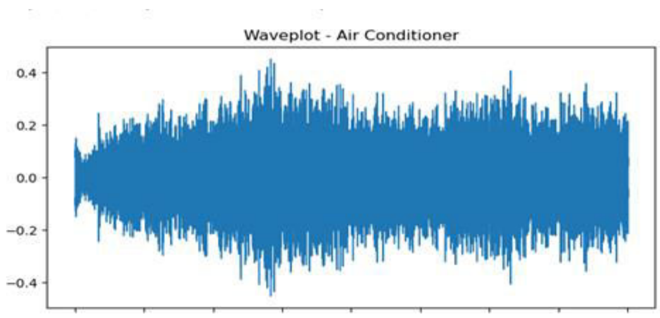


Fig.1. Linear-frequency power spectrogram



. Fig.2.Waveplot-Air Conditioner.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205348 |

This is the confusion matrix.

Fig.3.Confission matrix.

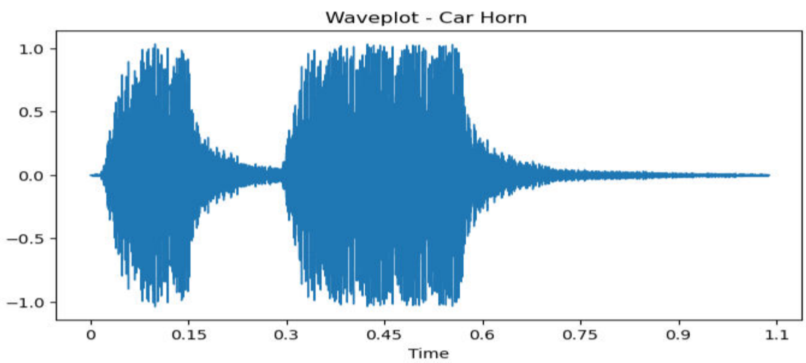


Fig.4.Waveplot-Car Horn

VI. CONCLUSION AND FUTURE WORK

This research presented a comprehensive study on environmental sound classification using Convolutional Neural Networks (CNNs). The primary objectives were to design and evaluate a CNN-based system capable of accurately classifying various environmental sounds. The study utilized the UrbanSound8K dataset to train and test the model, and the results demonstrated that the proposed CNN architecture is highly effective in classifying environmental sounds, achieving an accuracy of 96.7%.

The high performance of the model underscores the potential of deep learning techniques, particularly CNNs, in advancing the field of sound classification. This work contributes to the ongoing efforts to develop more accurate and reliable sound classification systems, which have numerous applications in areas such as surveillance, wildlife monitoring, and assistive technologies.

REFERENCES

[1] K. J. Piczak, “Environmental sound classification with convolutional neural networks,”in Proc. IEEE 25th Int. Workshop Mach. Learn. Signal Process., Sep. 2015, pp. 1–6.

[2] K. J. Piczak, “ESC: Dataset for environmental sound classification”in Proc. 23rdACM Int. Conf. Multimedia, 2015, pp. 1015–1018.

[3] D. Palzer and B. Hutchinson, “The tensor deep stacking network toolkit,” Proc. Int.Joint Conf. Neural Netw. (IJCNN), Jul. 2015, pp. 1–5

[4] Keras Documentation. [Online]. Available: https://keras.io

[5] SciPy: Open Source Scientific Tools for Python. [Online]. Available: https://www.scipy.org/

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205348 |

[6] F. Medhat, D. Chesmore and J. Robinson, "Masked conditional neural networks for sound classification", Applied soft computing, vol. 90, 2020

[7] K. Palanisamy, D. Singhania and A. Yao, "Rethinking CNN models for audio classification", November 2020.

[8] F. Medhat, D. Chesmore and J. Robinson, "Masked conditional neural networks for sound classification", Applied Soft Computing, vol. 90, 2020.

[9] Li David, Tam Jason and Toub Derek, "Auditory Scene Classification using Machine Learning Techniques", IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, 2021.

[10] Ye Jiaxing, Kobayashi Takumi and Murakawa Masahiro, "Urban Sound Event Classification based on Local and Global features Aggregation in Applied Acoustics", 2016.