Paper ID: 16

Title: South Asian Sounds: Audio Classification

Author(s) Name: Rajdeep Chatterjee, Pappu Bishwas, Sudip Chakrabarty, Tathagata Bandyopadhyay

Author(s) Affiliation: Kiit Deemed to be University, India, Technical University of Munich, Germany

Email : {cse.rajdeep, pappuovi8, sudipchakrabarty6, gata.tatha14}@gmail.com

Presenter : Sudip Chakrabarty

Outline

1. Introduction
2. Applications and Challenges
3. Motivation and Objectives
4. Related Work
5. Background Concepts
6. Proposed Framework
7. Experimental Set-up
8. Sample Dataset
9. Results and Analysis
10. Inference
11. Conclusion and Future Work

Introduction

Evolution:

- Shifted from manual methods like spectrogram analysis to machine learning.
- 1D-CNNs improved accuracy and speed in sound classification by applying MFCCs to classify South Asian urban sounds.

Applications and Challenges

- Essential for urban planning, noise control, and cultural preservation.
- Enables real-time monitoring of traffic and environmental noise.
- Identifies cultural sounds in diverse South Asian soundscapes.
- Conventional models struggle with complexity, emphasizing the need for efficient classification.

Motivation and Objectives

Motivation:

- Diverse South Asian environments require advanced sound classification.
- Real-time monitoring improves urban planning.
- Helps preserve and analyze traditional regional sounds.

Objective:

- Enhance sound classification with machine learning and deep learning.
- Achieve accurate categorization of diverse South Asian sounds.
- Address challenges of data complexity, noise, and sound diversity.
- Facilitate management of large-scale urban sound datasets.

Related Work

- Researchers have employed CNNs and LSTMs for urban sound classification [1].
- Signals have been transformed into images and classified using ResNet50v2 for sound analysis [2].
- Numerous studies have utilized MFCCs in combination with deep learning models for accurate sound classification.

Background Concepts

- Utilized MFCCs for feature extraction and a 1D-CNN for classifying urban sounds, capturing local patterns in sequential audio data.
- Provided Mel Spectrogram and waveform images for audio visualization, and used t-SNE to visualize high-dimensional feature clustering.
- Compared model performance against the UrbanSound8k dataset for benchmarking.
- Evaluated using a confusion matrix for classifications and accuracy/loss graphs to track model performance during training and testing.

Proposed Framework

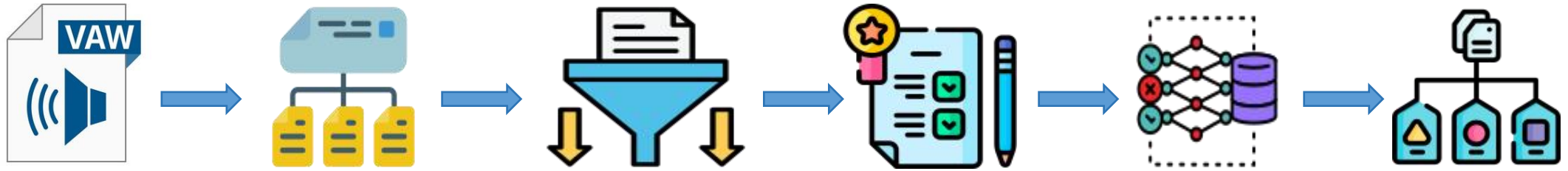

Fig.1 Overall workflow diagram

Experimental Set-up

- **Dataset:** The dataset, named "SAS-KIIT,"[3] contains audio recordings across various South Asian countries like India, Bangladesh, Sri Lanka, and Afghanistan, encompassing different sound categories.
- **Relevant Classes:** Focused on 5 relevant classes including Classroom Noise, Tanpura, Traditional Bengali Song, Traffic (Roadways), Train (Railways).
- **Data Size:** Comprises approximately 1.5 GB with 3750 audio files(WAV format) randomly distributed across ten folders, each up to 4 seconds long.

Experimental Set-up 2/2

- **Dataset:** Utilized the UrbanSound8k [4] dataset, containing urban sounds of a wide range of urban noises.
- **Relevant Classes:** The dataset includes 10 urban sound classes: Air Conditioner, Car Horn, Children Playing, Dog Bark, Drilling, Engine Idling, Gun Shot, Jackhammer, Siren, and Street Music.
- **Data Size:** Comprises approximately 8 GB of urban sounds, with 8,732 audio files(WAV format) distributed across ten folders, each up to 4 seconds long.

Sample Dataset

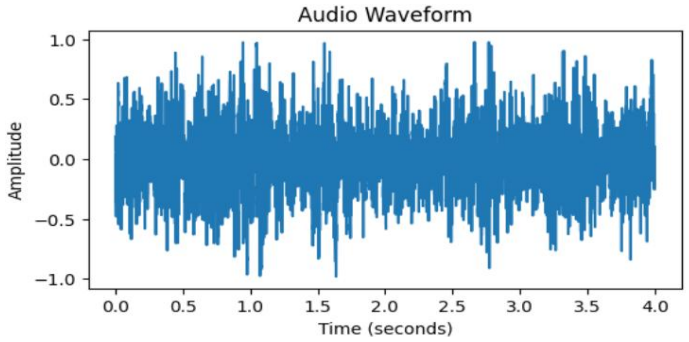

Fig 2: Train (Railways) Waveform

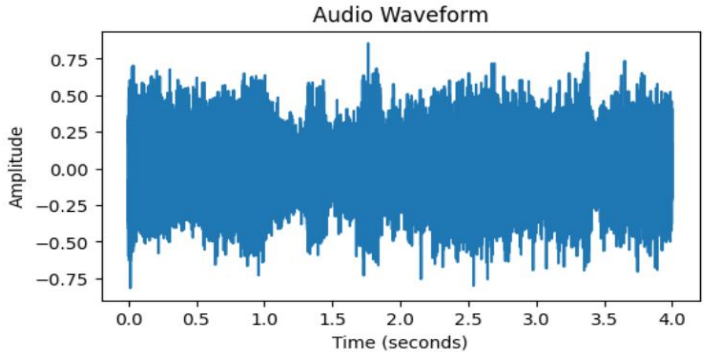

Fig 3: Traffic (Roadways) Waveform

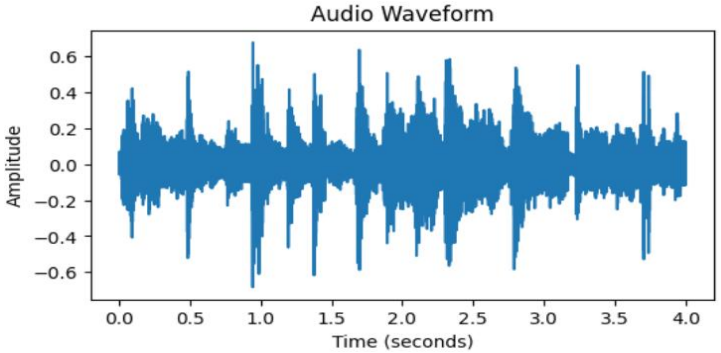

Fig 4 : Traditional Song Waveform

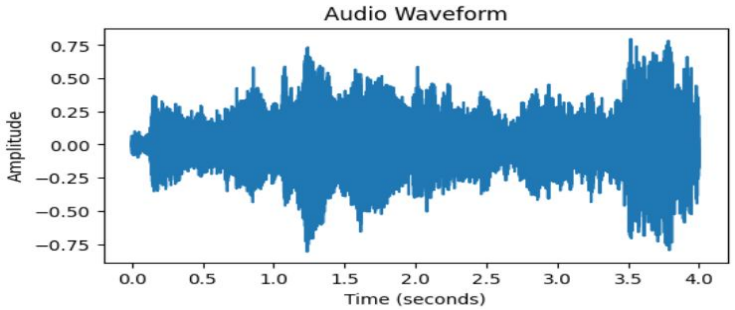

Fig 5: Classroom Noise Waveform

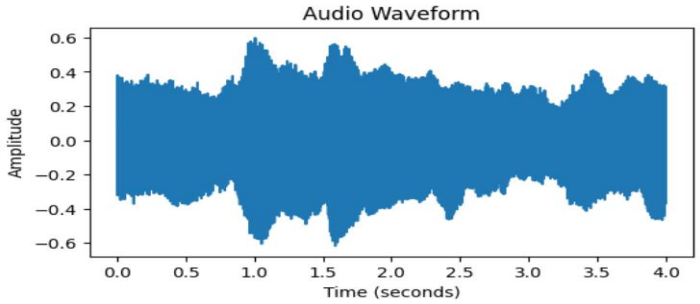

Fig 6: Tanpura Waveform

Results and Analysis

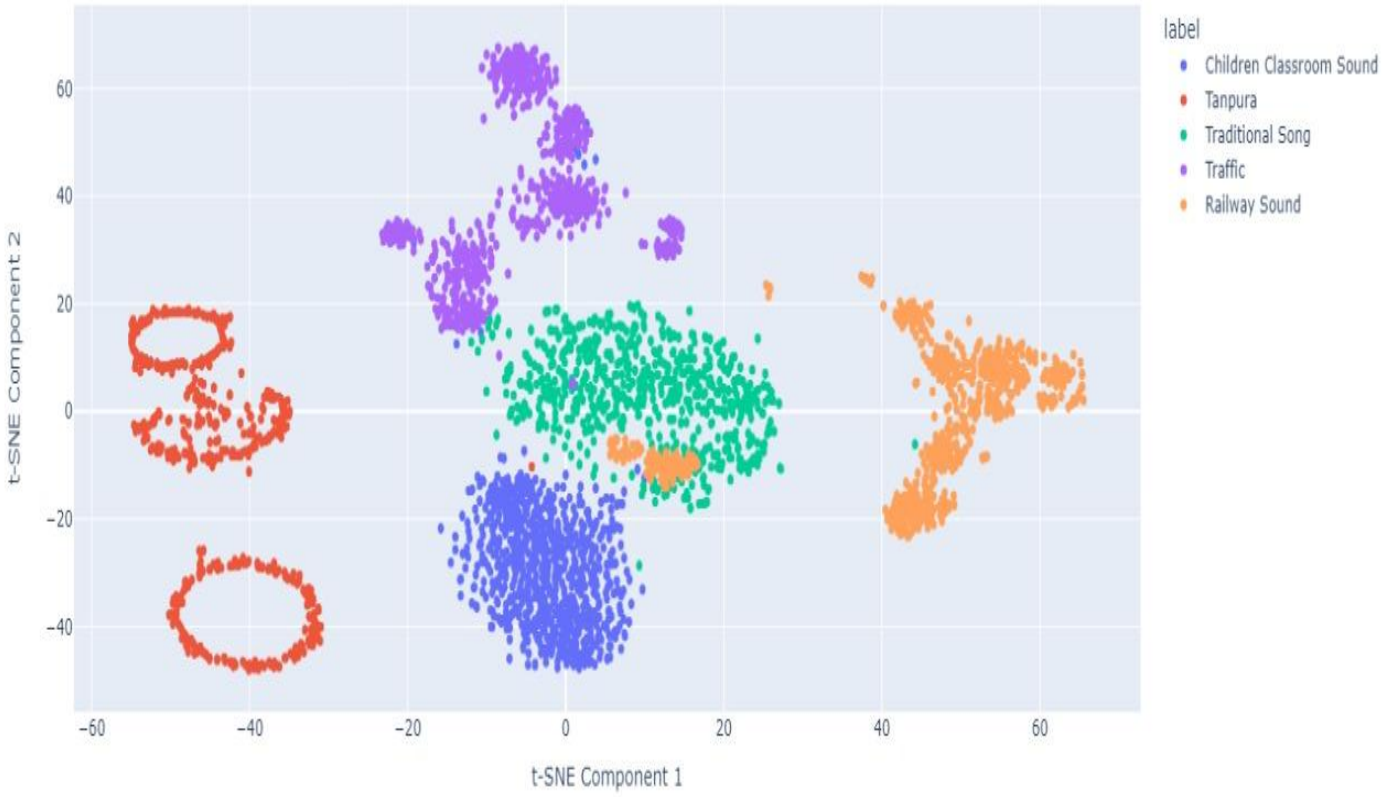

Fig 7: t-SNE Visualization

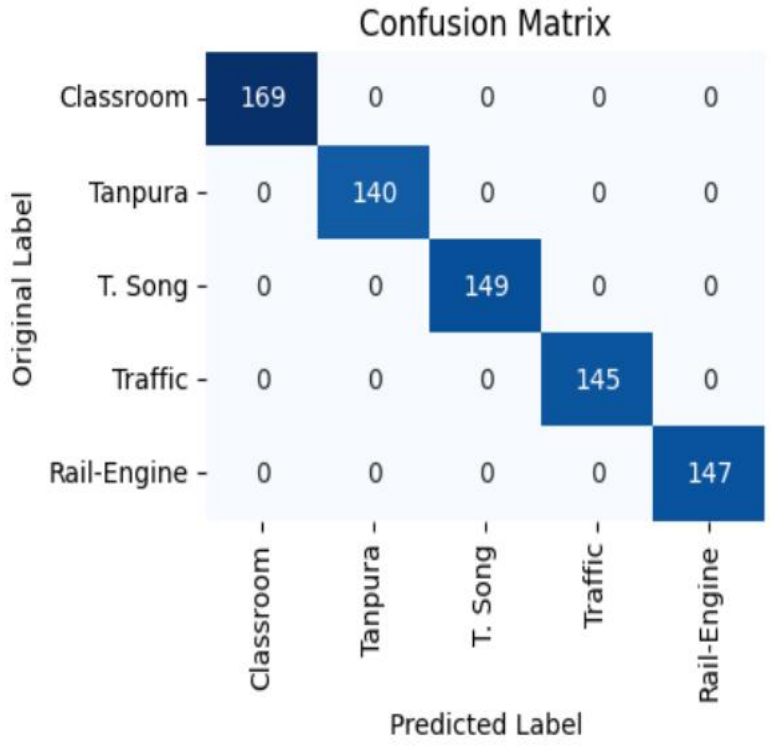

Fig 8: Confusion Matrix

Results and Analysis 2/2

Model	South Asian Sounds (SAS - KIIT)		UrbanSounds8k
	Accuracy (%)	Accuracy (%) 10-fold	Accuracy (%) 10-fold
1D-CNN (proposed SAS-CNN)	99.89	99.78 ± 0.20	94.26±0.20

- Utilized 10-fold cross-validation for balanced training and validation, enhancing model robustness.
- Applied early stopping and optimized step sizes to prevent overfitting and improve testing accuracy.

Inference

Performance (%) of Different Models on the UrbanSound8k Dataset.

Models	UrbanSounds8k
	Accuracy (%)
ResNet50V2 [5]	90.70
MLP [5]	82.11
DNN100 [1]	90.90
CNN [1]	87.15
LSTM [1]	90.15
1D-CNN (proposed SAS-CNN)	93.50

- Our model achieved the highest accuracy, surpassing all other approaches on the UrbanSound8k dataset.

Conclusion and Future Work

In summary:

- SAS-CNN (1D-CNN) outperforms state-of-the-art models in sound classification.
- Achieved high accuracy with holdout and 10-fold cross-validation.
- Demonstrated deep learning's potential for region-specific urban sound challenges.

In future:

- Expand dataset with more South Asian-specific sounds.
- Explore additional feature extraction techniques.
- Incorporate advanced classifiers and feature selection methods.
- Investigate real-time applications like environmental monitoring and smart city initiatives.

References

1. M. Bubashait and N. Hewahi, “Urban sound classification using dnn, cnn & lstm a comparative approach,” in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). IEEE, 2021, pp. 46–50.

2. D. Xv and L. Yang, “Research on urban audio classification based on residual neural network,” in 2021 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2021, pp. 200–203.

3. “SAS-KIIT”, https://on.soundcloud.com/HT5SXWJQg2jgGzvX7

4. “Urbansound8k dataset,” https://urbansounddataset.weebly.com/urbansound8k.html, July 2014, accessed: 2024-1-06.

5. D. Xv and L. Yang, “Research on urban audio classification based onresidual neural network,” in 2021 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2021, pp. 200–203.

Thank You

Dr. Rajdeep Chatterjee
KIIT Deemed to be University, Bhubaneswar-751024, India

Sudip Chakrabarty
KIIT Deemed to be University, Bhubaneswar-751024, India

Pappu Bishwas
KIIT Deemed to be University, Bhubaneswar-751024, India

Tathagata Bandyopadhyay
Technical University of Munich, Germany