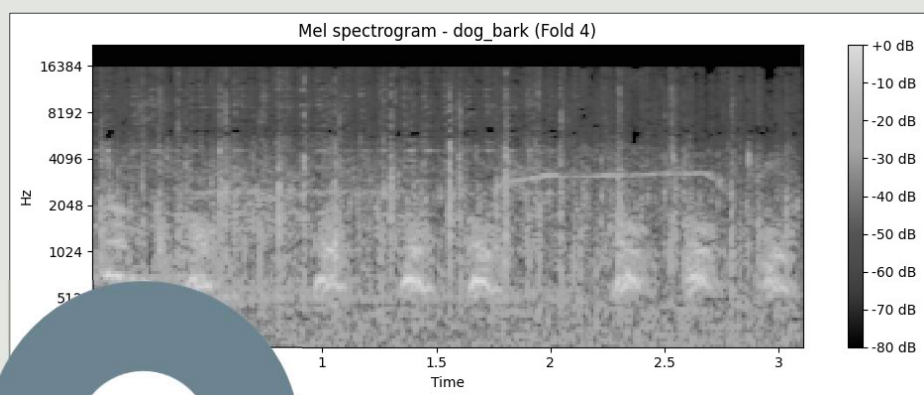


ANNA ESE PROJECT

ENVIRONMENTAL SOUND CLASSIFICATION

USING DEEP CNNs AND AUGMENTED SPECTROGRAMS



Group Members

Kaustubh Wagh
Nikita Kadam
Apoorva Singh
Yash Sawant

202201070021
202201070024
202201070030
202201070032

Contents

• Project Overview	3
• Data Augmentation	4
• CNN Architecture	5
• Training Setup	6
• Results & Performance	7
• Model Valuation	8
• Conclusion & Future Scope	9

Project Overview

Objective: Automatically classify urban sounds into 10 categories



Technologies: PyTorch, Librosa, Audiomentations, Scikit-learn

Sound Classes

- Air Conditioner,
- Car Horn,
- Children Playing,
- Dog Bark,
- Drilling,
- Engine Idling,
- Gun Shot,
- Jackhammer,
- Siren,
- Street Music

Data Augmentation

Augmentation Techniques :

- Gaussian Noise: 0.001-0.015 amplitude
- Time Stretch: 0.8x - 1.25x speed
- Pitch Shift: ± 4 semitones
- Time Shift: ± 0.5 seconds

Benefits: Improved generalization, reduced overfitting, real-world variation simulation

CNN Architecture

Model Structure: Input (1×128×173) → 3 Conv Blocks → Global Pool → FC Layer

Conv Blocks:

- Block 1: 32 filters (3×3) + BatchNorm + ReLU + MaxPool
- Block 2: 64 filters (3×3) + BatchNorm + ReLU + MaxPool
- Block 3: 128 filters (3×3) + BatchNorm + ReLU + MaxPool

Key Features: Global Average Pooling, Dropout (0.5), Progressive filter increase

Training Setup

- Optimizer: Adam (lr=0.001)
- Batch Size: 16, Max Epochs: 80
- Loss: CrossEntropyLoss

Early Stopping:

- Patience: 10-15 epochs
- Monitor: Validation accuracy
- Auto best weight restoration

Data Split: 80% Train | 10% Validation | 10% Test

Results & Performance

Key Results:

- Test Accuracy: 85-90%
- Convergence: 15-25 epochs
- Effective overfitting control
- Consistent performance

Success Factors:

- Mel spectrograms provide excellent features
- Data augmentation improves robustness
- Early stopping prevents overfitting
- Progressive CNN captures hierarchical patterns

Model Evaluation

Evaluation Methods:

- Classification Report (Precision, Recall, F1-Score)
- Confusion Matrix visualization
- Training curves analysis
- Sample prediction visualization

Technical Implementation:

- GPU/CPU auto-detection
- Comprehensive metric calculation
- Visual performance analysis
- Per-class performance breakdown

Conclusion & Future Scope

Achievements:

- End-to-end audio classification system
- 85-90% accuracy on urban sounds
- Robust preprocessing pipeline
- Comprehensive evaluation framework

Future Improvements:

- Architecture: Attention mechanisms, Transformers
- Features: Multi-scale spectrograms, feature fusion
- Data: Advanced augmentation, larger datasets
- Deployment: Real-time inference, mobile optimization

THANK YOU