Gunshot Detection System On Edge Devices

Under the guidance of **Dr. SHAIK RIYAZ HUSSAIN SIR**
Prepared by:
P. Praveen Kumar - N210402

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOGIES, NUZVID

25 October 2025

Table of Contents

Introduction

Gunshot Detection — A Need for Intelligent Safety Systems

Public safety systems increasingly rely on AI-driven solutions for real-time threat detection. Gunshot detection is critical for:

- Rapid emergency response in urban and public spaces.
- Accurate identification of firearm-related acoustic events.
- Integration with surveillance networks for automated alert systems.

Challenge: Traditional sound-based systems often fail in noisy or cluttered environments.

Why Gunshot Detection on Edge Devices?

Cloud-based gunshot detection systems face latency and privacy challenges. To overcome these, lightweight models on embedded platforms like **Raspberry Pi** are ideal.

- Enables **real-time inference** close to the source.
- Reduces dependency on high-speed internet.
- Supports **scalable and low-power deployment**.
- Enhances community safety through rapid local alerts.

Goal: Design an efficient and accurate AI-based gunshot detection model for edge deployment.

Base Paper Overview

Real-time Gunshot Detection System

The base paper proposes a real-time gunshot detection system for enhanced public safety. It captures environmental audio, extracts features (MFCC or YAMNet), and uses neural network models for accurate gunshot classification suitable for edge devices like Raspberry Pi.

- Audio capture and feature extraction
- ML-based gunshot classification (up to 96% accuracy)
- Optimized for real-time, low-power hardware

Analysis of Base Paper

The base paper made significant contributions to real-time gunshot detection but also had limitations that inspired our enhancements:

- **Contributions:**
  - Evaluation of multiple hardware platforms for real-time inference feasibility.
  - Consideration of noisy environments and confounding audio events.
- **Limitations:**
  - Hardware optimization and low-latency inference were limited.
  - Model compression techniques like pruning or quantization were not explored.
  - Lightweight embedding-only inference for resource-constrained devices was not implemented.
  - Model was trained on small dataset only leading to overfitttig.

Gunshot Detection System Overview

The paper presents a real-time gunshot detection system integrated with camera surveillance. Audio data is pre-processed to improve model accuracy.

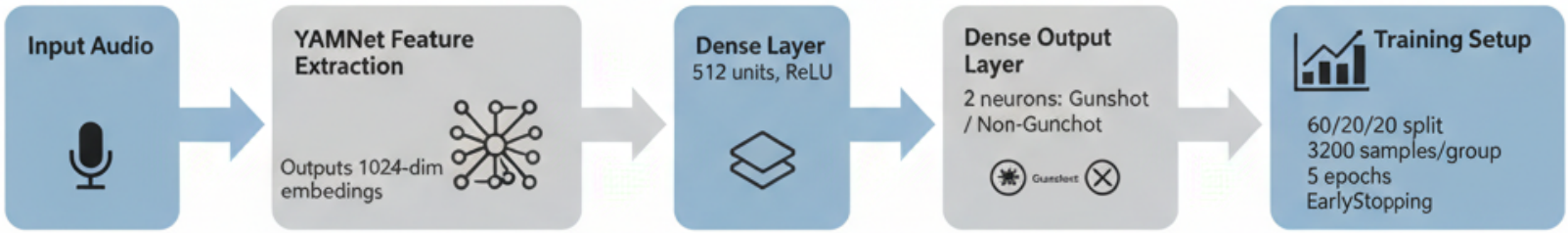
- Gunshot sounds from various firearms (AK-47, MP5, M16, etc.), resampled to 1-sec 22050 Hz and filtered for low-energy noise, yielding 3210 samples.
- Non-gunshot sounds (thunder, fireworks, drums, doors, clapping, barks) collected from YouTube, totaling 7758 samples after preprocessing.

Modeling Approaches

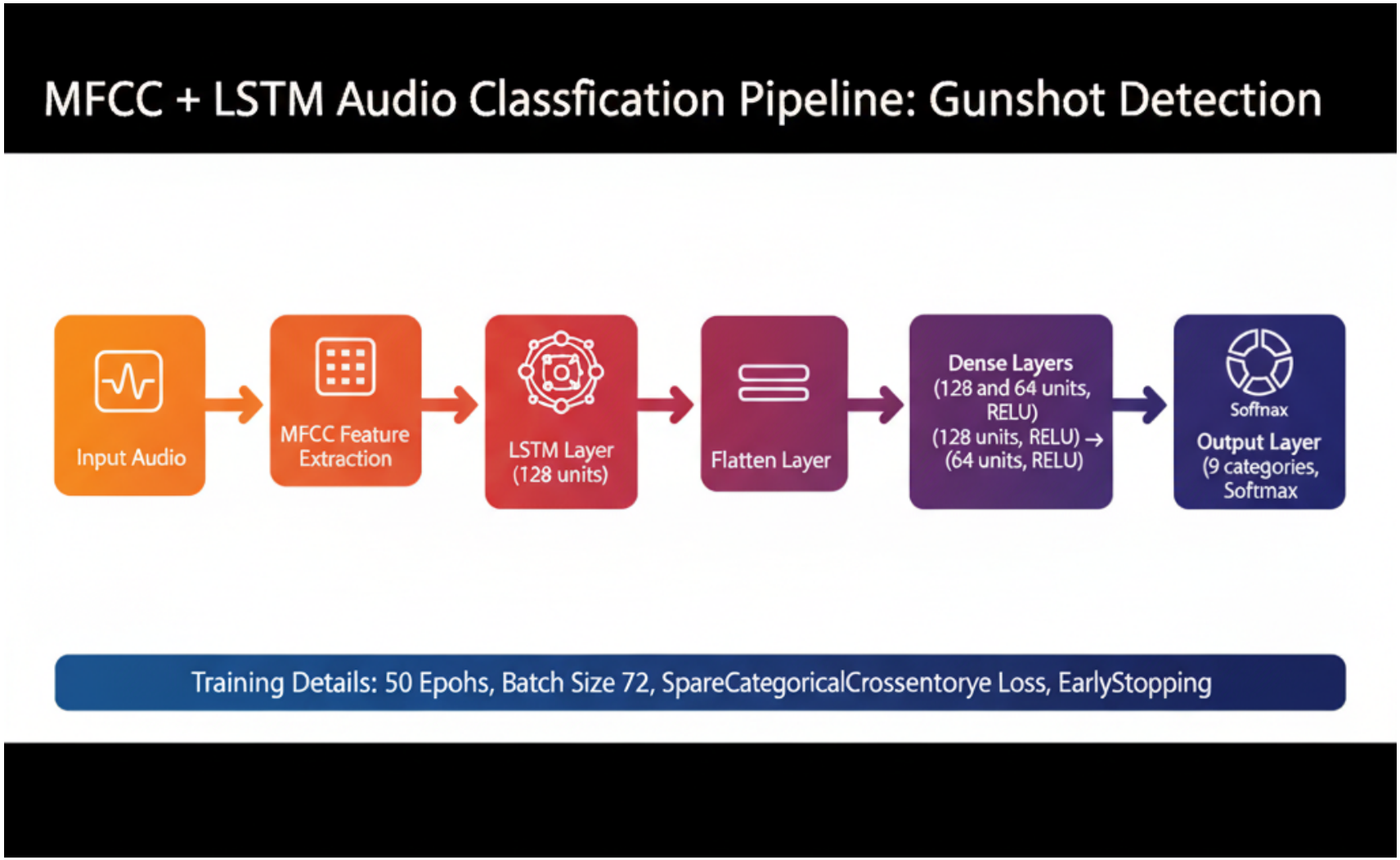
Two ML approaches were used for gunshot detection:

1. **YAMNet Transfer Learning:**
   - 1024-dim YAMNet embeddings as input.
   - Three-layer network (512-unit dense + 2-output).
   - 60/20/20 train/val/test split, 3200 samples/group.
   - EarlyStopping, trained for 5 epochs.
2. **MFCC-Based LSTM:**
   - Audio $\rightarrow$ MFCC features.
   - LSTM: 128-unit LSTM $\rightarrow$ Flatten $\rightarrow$ Dense layers (128, 64) $\rightarrow$ 9-output.
   - 50 epochs, batch size 72, SparseCategoricalCrossentropy loss, EarlyStopping.

YAMNet Transfer Learning Flow



MFCC-Based LSTM Flow



Gunshot Detection Model Performance

A. TensorFlow YAMNet Transfer Learning Model:

- Test set: Loss = 0.0490, Accuracy = 98.75%
- Confusion Matrix (1280 samples): **Gunshot: 637/640 correct**, **Non-gunshot: 635/640 correct**
- Minimal misclassification demonstrates strong predictive power for real-world deployment.

B. MFCC-Based LSTM Model:

- Test set: Loss = 0.1676, Accuracy = 96.95%
- Confusion Matrix (1280 samples): **Gunshot: 623/640 correct**, **Non-gunshot: 618/640 correct**
- High accuracy indicates strong generalization for real-world use.

Results Obtained

Test Performance Comparison

Both models achieved high accuracy on training, validation, and test sets. Final evaluation results are as follows:

YAMNet Transfer Learning

- Test Loss: 0.0384
- Test Accuracy: 98.52%
- Confusion Matrix:

$$\begin{bmatrix} 636 & 4 \\ 10 & 630 \end{bmatrix}$$

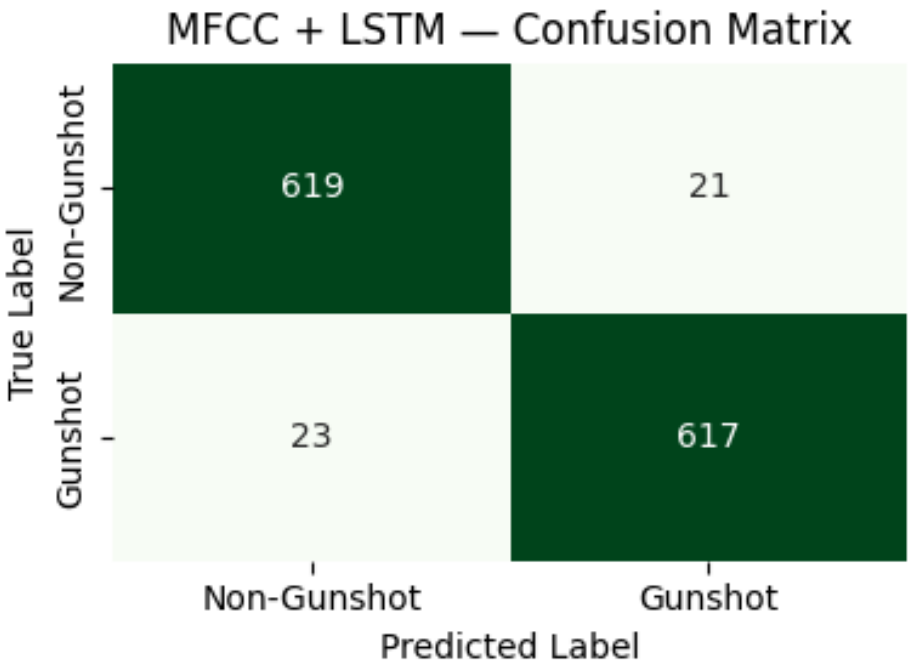
MFCC-Based LSTM

- Test Loss: 0.1706 Test
- Accuracy: 96.94%
- Confusion Matrix:

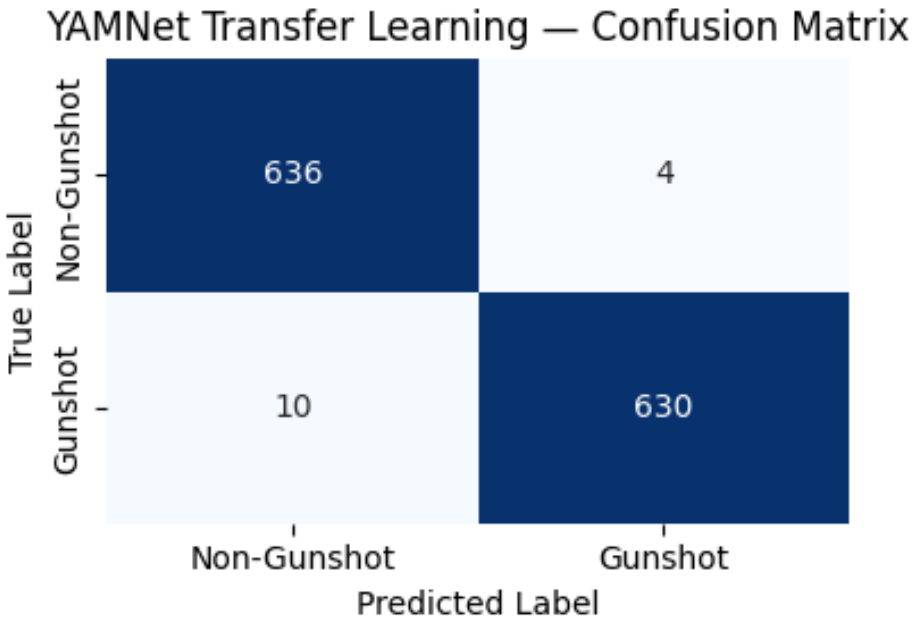
$$\begin{bmatrix} 619 & 21 \\ 23 & 617 \end{bmatrix}$$

MFCC vs YAMNet Feature Matrices

MFCC_Matrix



YAMNet_Matrix



Spectral–temporal representation extracted using MFCC features.

Deep audio embeddings learned by YAMNet pretrained model.

The MFCC matrix captures handcrafted frequency features, while YAMNet embeddings provide high-level semantic representations learned from large-scale audio datasets.

YAMNet vs MFCC-LSTM Evaluation Summary

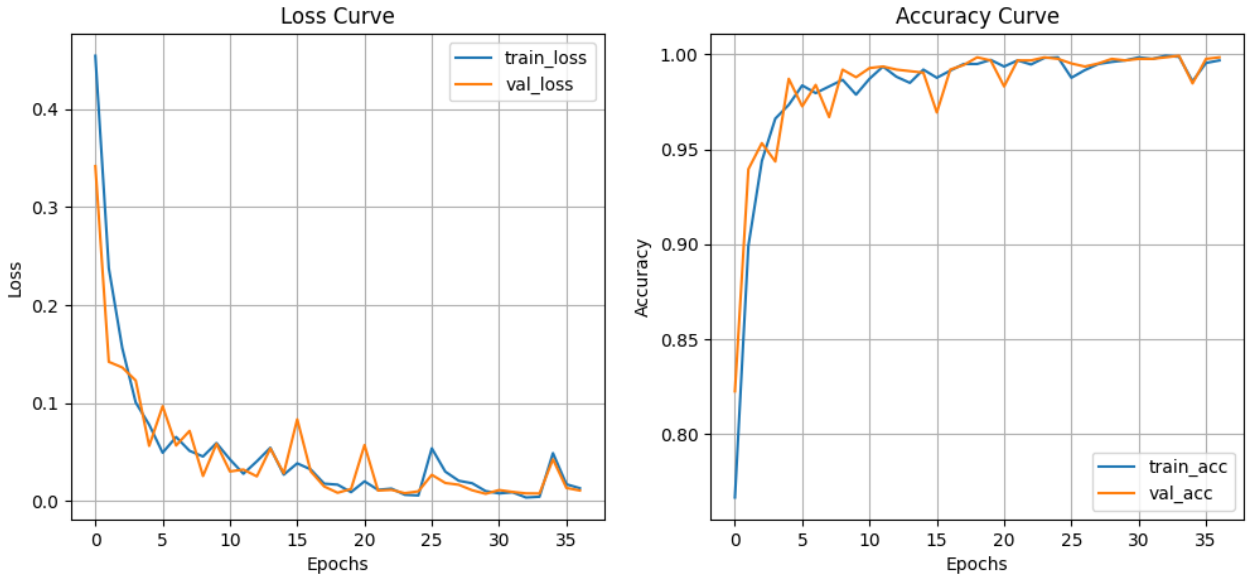
Metric	YAMNet Transfer Learning	MFCC-Based LSTM
Test Loss	0.0384	0.1706
Test Accuracy	98.52%	96.94%
Gunshot Correct (out of 640)	636	619
Non-Gunshot Correct (out of 640)	630	617
Total Misclassifications	14	44
Inference Complexity	Low (Dense only)	Moderate (LSTM + Dense)
Deployment Suitability	**High**	Medium

Conclusion: YAMNet-based model achieves higher accuracy and efficiency for real-time deployment.

MFCC + LSTM Model Performance

The MFCC-LSTM model achieved strong accuracy, though with slightly higher loss and slower convergence.

Accuracy and Loss Curves



Final Test Accuracy: 96.94% — **Test Loss:** 0.1706

Model Advancements

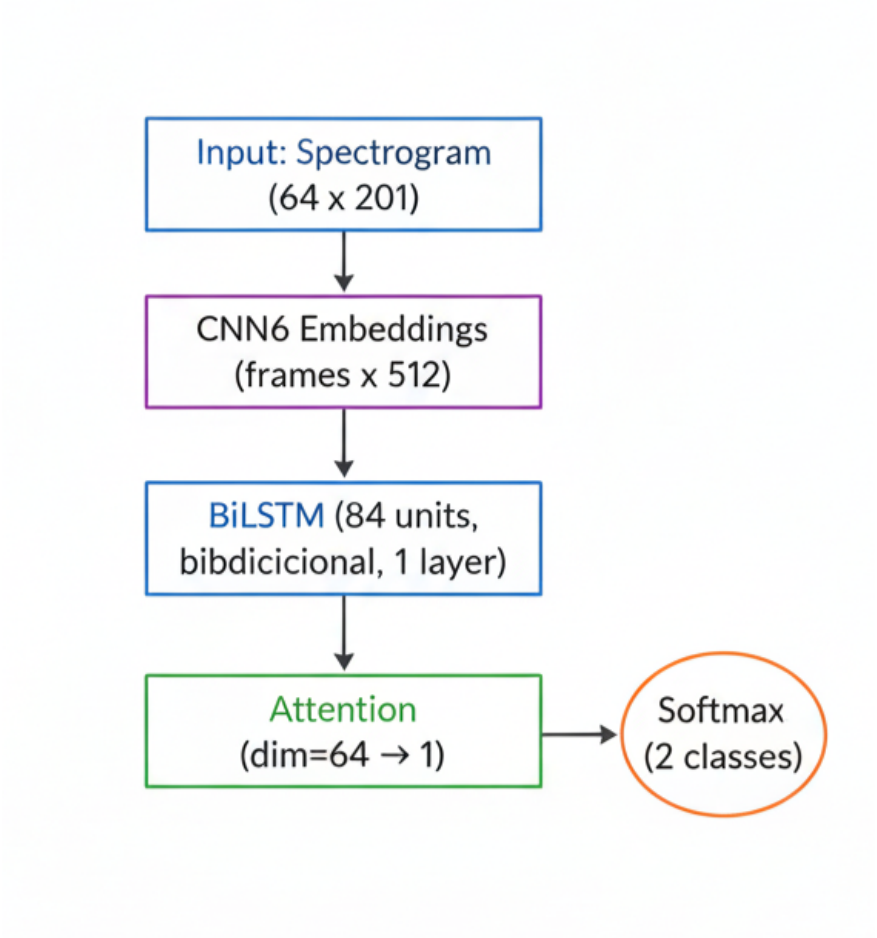
Model Architecture Overview

Goal: Enhance gunshot detection by combining spectral–temporal learning with attention mechanisms.

Model Highlights:

- **CNN14 (PANNs):** Extracts deep log-Mel spectral embeddings.
- **BiLSTM:** Captures forward–backward temporal dependencies.
- **Attention Layer:** Focuses on key temporal frames, suppresses silence or background noise.
- **Dense Head:** Fully connected layers with 128 and 64 ReLU units + dropout for regularization.
- **Output Layer:** Softmax activation producing Gunshot / Non-Gunshot classification.

Proposed Model Architecture



CNN14 → BiLSTM → Attention → Dense → Output

References

References I

[1] Xinzhang Xiong, "Real-time Gunshot Detection System Integration to Camera Surveillance System", Pennsylvania State University, State College, USA. Email: xpx5059@psu.edu

[2] Google Research, *"YAMNet: Pretrained Audio Event Classification Network"*, Available at: `https://github.com/tensorflow/models/tree/master/research/audioset/yamnet`

[3] Kong, Qiuqiang, et al. *"PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition"*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020.

[4] T. Giannakopoulos, *"Audio Feature Extraction using MFCCs for Environmental Sound Classification"*, Elsevier, 2015.

[5] S. Han, H. Mao, and W. J. Dally, *"Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding"*, ICLR, 2016.

Thank you!