ENGINEERING PRACTICE REPORT

Development of an Acoustic Discrimination Device for Ambulance and Police Vehicle Sirens Using Artificial Intelligence on an Embedded Device

On-Edge Audio Classification Based on Mel-Filterbank Energy and a 1-Dimensional Convolutional Neural Network

Dwi Ahmad Dzulhijjah
Student ID 2202411012
Engineering Professional Program (PSPPI)
Institut Teknologi Indonesia · 2024

Abstract—In traffic management and emergency response, the ability to distinguish ambulance and police vehicle sirens quickly and accurately is decisive for response time. This report documents the development of an AI-based siren discrimination device that runs directly on an Arduino Nano 33 BLE Sense embedded device. Audio is captured by an MP34DT05 microphone, its features extracted with Mel-Filterbank Energy (MFE), then classified by a one-dimensional convolutional neural network into three classes: ambulance siren, police siren, and ambient noise. The model was trained on the Edge Impulse platform and optimized with the EON Optimizer to fit the device's power and memory constraints. The on-device inference result directly issues a traffic-light priority signal for emergency vehicles, with no dependency on a server or internet connection on the critical path. On the validation set the model reached 96.5% accuracy with a weighted F1 of 0.97 and an area under the ROC curve of 1.00, at 9 ms inference per window on the embedded device. This report presents the methodology, system architecture, evaluation results, and recommendations for further development.

Index Terms—siren discrimination, audio classification, Mel-Filterbank Energy, convolutional neural network, edge computing, Edge Impulse, emergency response.

I. INTRODUCTION

1.1 Background

The ability to respond to emergencies quickly and accurately is essential in modern transportation systems. One critical aspect is detecting and distinguishing the sirens of emergency vehicles such as ambulances and police cars. A failure to discriminate correctly can cause response delays that lead to life-threatening situations. A technology that can automatically and efficiently distinguish these sirens is therefore needed.

Advances in artificial intelligence, particularly when applied to embedded devices, offer real-time data processing with the efficiency and speed required in emergencies. Developing this device draws on several engineering disciplines: electrical engineering, computer engineering, and industrial engineering. The system is designed with reference to the Indonesian National Standard (SNI) for intelligent transportation systems and to guidance from BPPT and the Ministry of Transportation, and is consistent with the national mid-term development agenda on the use of information technology in transportation and health.

1.2 Objective

The objective of this engineering practice is to develop an acoustic discrimination device for ambulance and police sirens using artificial intelligence implemented on an embedded device, in order to improve traffic management and emergency response through fast and accurate siren detection.

1.3 Scope

- Collecting emergency-vehicle siren audio from various sources and situations.
- Developing and training an AI model using the collected dataset.
- Implementing the model on the Arduino Nano 33 BLE Sense embedded device for real-time detection.
- Testing and evaluating the device's performance under various environmental conditions.
- Preparing a report covering the methodology, results, and recommendations for improvement.

1.4 Implementation Method

Implementation combines two frameworks. The **System Development Life Cycle (SDLC)** structures the systems engineering through planning, analysis, design, implementation, testing, and maintenance. **CRISP-DM (Cross-Industry Standard Process for Data Mining)** structures the data workflow through business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

II. PROBLEM FORMULATION AND SYSTEM ARCHITECTURE

2.1 Problem Formulation

The core problem is the difficulty of distinguishing emergency-vehicle sirens under varied environmental conditions, which can cause response delays. Additional technical challenges include the limited processing power of embedded devices and the need to maintain high performance across varying acoustic conditions.

2.2 Processing Chain

The system is arranged as an edge processing chain. The MP34DT05 microphone built into the Arduino Nano 33 BLE Sense captures ambient audio. The signal's features are extracted with Mel-Filterbank Energy, then fed into a neural network trained to classify the audio type. Inference runs directly on the microcontroller in C++. The on-device classification result directly turns the traffic light green to give right of way to the emergency vehicle. The output is also shown on a serial monitor.

Figure 2.1 Block diagram of the edge processing chain, from audio acquisition to traffic-light actuation.

2.3 Hardware, Software, and Algorithm

- **Hardware:** Arduino Nano 33 BLE Sense (Cortex-M4F 64 MHz) with built-in MP34DT05 microphone; a traffic light as the actuator; a serial interface for monitoring.
- **Software:** Edge Impulse for model development; the Arduino IDE for firmware; the EON Optimizer for memory efficiency; a command-line interface for inference testing.
- **Algorithm:** Mel-Filterbank Energy feature extraction; a one-dimensional convolutional neural network for classification; C++ inference on the device.

2.4 Standards and Constraints

Development refers to the SNI for intelligent transportation systems and to guidance from BPPT and the Ministry of Transportation. The main constraints are the limited compute power of the embedded device and the wide variation of environmental conditions.

III. MODEL CONFIGURATION AND TRAINING

3.1 Signal Processing Parameters (MFE)

Audio is sampled at 16 kHz with a 1,000 ms window and a 500 ms stride. MFE extraction uses a 0.02 s frame length, 0.01 s frame stride, 40 mel filters, an FFT length of 256, a low frequency of 0 Hz, and a noise-floor normalization of –52 dB. This configuration yields 3,960 input features represented as a 40-column mel spectrogram.

Parameter	Value	Parameter	Value
Sample frequency	16,000 Hz	Mel filter count	40
Window size	1,000 ms	FFT length	256
Window stride	500 ms	Low frequency	0 Hz
Frame length	0.02 s	Noise floor	–52 dB
Frame stride	0.01 s	Input features	3,960

Table 3.1 Mel-Filterbank Energy extraction parameters.

3.2 Network Architecture and Training

The network takes 3,960 features, reshapes them into 40 columns, then passes them through two one-dimensional convolution blocks (8 and 16 filters, kernel size 3) each followed by 0.25 dropout, a flatten layer, and a three-class output layer. Training ran for 100 cycles with a learning rate of 0.005. Of 1,406 data windows with a balanced class composition, 1,124 windows were used for training and 282 for validation. Convergence was fast: validation accuracy passed 91% around epoch 10 and settled in the mid-90s with a peak of 96.8%, with one brief spike at epoch 61 that recovered within two epochs.

IV. RESULTS AND EVALUATION

Evaluation on the validation set shows high, balanced performance. Overall accuracy reached 96.5% with a loss of 0.08. After balancing the class composition, the ambulance class is detected perfectly and never confused with police; the remaining errors lie on the boundary between police sirens and noise.

Metric	Value	On-device performance	Value
Accuracy (validation)	96.5%	Inference time	9 ms
Loss	0.08	Peak RAM usage	14.7 KB
Area under ROC	1.00	Flash usage	40.9 KB
Weighted precision	0.97	Inference engine	EON Compiler
Weighted recall / F1	0.96 / 0.97	Quantization	int8

Table 4.1 Summary of model performance metrics and the on-device execution profile.

Confusion Matrix	Ambulance	Police	Noise
Ambulance	100%	0%	0%
Police	0%	94.4%	5.6%
Noise	0%	9.1%	90.9%
F1	1.00	0.96	0.86

Table 4.2 Confusion matrix on the validation set (rows = true class). Ambulance is detected perfectly; errors concentrate on the police-noise boundary.

V. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

- The AI model was successfully implemented on an embedded device and can distinguish ambulance and police sirens with high accuracy (96.5% on validation, with perfect ambulance detection).
- The system operates efficiently and reliably with 9 ms inference and a small memory footprint (14.7 KB RAM, 40.9 KB flash), within the constraints of the Arduino Nano 33 BLE Sense (Cortex-M4F).
- With accurate siren discrimination, the system can potentially improve traffic management and emergency response, supporting public safety.

5.2 Recommendations

- Expand the dataset with more diverse environmental conditions to improve robustness.
- Apply audio data augmentation and further optimization to reduce noise-class errors.
- Conduct larger-scale testing under varied real-world conditions.
- Add detection of other emergency sound types (e.g. fire trucks) and integrate with traffic management systems.
- Publish the results and strengthen collaboration with relevant institutions.

References

[1] Badan Pengkajian dan Penerapan Teknologi (BPPT). (2020). Indonesian National Standard (SNI) for Intelligent Transportation Systems. Jakarta: BPPT Press.

[2] Ministry of Transportation of the Republic of Indonesia. (2020). Regulation of the Minister of Transportation on the Use of Information Technology in Transportation Systems. Jakarta.

[3] Mujumdar, S. & Thombare, M. V. (2018). Artificial Intelligence in Real-Time Sound Classification for Emergency Vehicles. Journal of Signal Processing Systems, 91(2), 159-172.

[4] Tan, J., Wang, J., & Zhang, Z. (2019). Edge Computing for Emergency Response: Leveraging AI and IoT for Situational Awareness. IEEE Internet of Things Journal, 6(3), 5201-5211.

[5] Zhao, Q., Huang, C., & Wu, Y. (2021). Real-Time Siren Detection Using Convolutional Neural Networks on Embedded Devices. IEEE Transactions on Vehicular Technology, 70(4), 3712-3721.