TRANSPORT PROBLEMS 2025 Volume 20 Issue 4
PROBLEMY TRANSPORTU DOI: 10.20858/tp.2025.20.4.09

Keywords: embedded artificial intelligence; unmanned aerial vehicle; railway infrastructure health monitoring; STM32; audio classification; wireless communication; real-time threat detection

Dmytro BOSYI[1]*, Oleg SABLIN[2], Iryna POTAPCHUK[3], Andrii USENKO[4]

EMBEDDED AI FOR AUDIO-BASED DRONE DETECTION IN CRITICAL RAILWAY INFRASTRUCTURE

Summary. With the increasing threat of unmanned aerial vehicles to critical railway infrastructure, the need for advanced detection technologies has become more urgent. This paper reviews existing railway monitoring solutions and outlines their limitations in identifying aerial threats. An acoustic analysis is conducted to extract distinctive unmanned aerial vehicle sound patterns using Mel-frequency cepstral coefficients, which serve as primary features for classification. Neural network models are applied to detect and differentiate aerial threats from environmental noise, achieving high recognition accuracy. The study also describes the development of an embedded artificial intelligence system based on STM32 microcontrollers, which combines real-time digital signal processing with efficient on-device neural inference. This solution offers a scalable and energy-efficient platform for decentralized audio-based drone detection in railway security applications.

1. INTRODUCTION

Both electrified and non-electrified railway infrastructures are critical to the efficient functioning of transportation networks across the globe. Electrified railways offer substantial environmental benefits, as electric trains produce few greenhouse gas emissions, reduce air pollution, and contribute to sustainability goals. Compared to diesel-powered systems, they are more energy-efficient, cost-effective, faster, accelerate more quickly, and generate less noise pollution. These advantages support sustainable transportation solutions while improving operational efficiency and the quality of life of people living near rail corridors. Countries worldwide are investing heavily in electrification, with regions like Europe, Asia, and North America leading the way, recognizing the immense benefits of clean, fast, and reliable transport networks.

However, electrified and non-electrified railway systems face significant challenges related to maintenance, safety, and operational reliability. Overhead lines in electrified networks are particularly prone to wear and tear, weather-related damage, and require frequent maintenance to ensure safe and efficient operation. These challenges can lead to service disruptions, high maintenance costs, and increased downtime. In this context, the Artificial Intelligence of Things (AIoT) - a combination of the Internet of Things (IoT) and artificial intelligence (AI) - provides promising solutions for monitoring infrastructure in real time. While the IoT enables the collection of critical data, such as voltage and current across railway lines, AI algorithms can analyze these data streams, detect anomalies, and predict potential failures. This integration significantly increases the efficiency of online detection and

[1] Ukrainian State University of Science and Technologies; 49010, Lazaryana 2, Dnipro, Ukraine; e-mail: d.o.bosyi@ust.edu.ua; orcid.org/0000-0003-1818-2490
[2] Ukrainian State University of Science and Technologies; 49010, Lazaryana 2, Dnipro, Ukraine; e-mail: o.i.sablin@ust.edu.ua; orcid.org/0000-0001-6784-648X
[3] Ukrainian State University of Science and Technologies; 49010, Lazaryana 2, Dnipro, Ukraine; e-mail: i.y.potapchuk@ust.edu.ua; orcid.org/0000-0002-5985-1040
[4] Ukrainian State University of Science and Technologies; 49600, Nauky 4 av., Dnipro, Ukraine; e-mail: a.u.usenko@ust.edu.ua; orcid.org/0000-0001-7467-6220
* Corresponding author. E-mail: d.o.bosyi@ust.edu.ua

monitoring, which enables predictive maintenance that anticipates failures, optimizes maintenance schedules, reduces energy losses, and ensures that the railway system operates safely and efficiently.

In recent years, low-cost unmanned aerial vehicles (UAVs), commonly known also as drones, have presented a new threat to railway infrastructure. The ongoing conflict in Ukraine has highlighted the vulnerability of critical infrastructure, with drones being used to target and damage valuable assets like railway and energy systems. Traditional air-defense systems struggle to detect and intercept these small, low-flying devices, which can be difficult to track using conventional methods. This has created an urgent need for innovative approaches to drone detection and identification within infrastructure monitoring systems.

This paper addresses these challenges by exploring how audio-based drone detection systems can be integrated into real-time railway infrastructure monitoring. Specifically, it focuses on leveraging the capabilities of wide area monitoring systems to monitor the mechanical parameters of railway assets, such as the tension and wear on overhead lines. Moreover, it investigates how these systems can be enhanced by incorporating decentralized, autonomous AI-driven solutions for UAV detection using acoustic analysis to identify the unique sound signatures of drones.

The authors aimed to comprehensively review the current state of railway monitoring technologies, including the integration of the AIoT for infrastructure health monitoring and drone threat detection. This research work:

1. Analyzes the current state of existing monitoring solutions for railway infrastructure.
2. Investigates the acoustic signatures of aerial threats and their identification.
3. Explores methods for integrating machine learning and neural networks into real-time drone detection.
4. Develops a hardware and software solution for embedded AI-driven audio-based drone detection, optimized for use on microcontroller platforms such as STM32.

By addressing these objectives, this study aims to contribute to the development of more resilient and secure railway systems capable of protecting critical assets from emerging threats while maintaining operational efficiency.

2. EXISTING SOLUTIONS FOR RAILWAY INFRASTRUCTURE MONITORING

Numerous studies [1-2] have addressed the monitoring and diagnostics of railway infrastructure to improve operational safety and efficiency. One such work introduces the PAMAR SMD system − a diagnostic device for railway contact networks that collects real-time data on supply voltage, temperature, humidity, and overall system status. This system enhances safety, reduces restoration costs, and minimizes delay-related compensations. Another study monitored current collectors during actual operation, presenting a station-based approach that detects issues by analyzing the vertical displacement of contact wires caused by passing trains. By identifying excessive or insufficient contact forces, this system ensures safe and reliable pantograph operation.

The Siemens Sicat® Catenary Monitoring System (CMS) [3] offers the continuous and contactless monitoring of tensile forces in overhead wires. It evaluates sensor data and transmits filtered information to control centers for rapid fault detection and localization. Notable features include non-invasive measurements that avoid wear on mechanical components, reliable detection through swing lever inclination tracking, easy retrofitting for existing systems, and low maintenance demands. Event-driven alarms and targeted fault messaging further enhance operational efficiency without data overload.

Sensonic GmbH [4] has developed the SonicTwin® system, which employs fiber-optic vibration sensing to deliver predictive insights into track integrity and potential safety risks. The solution provides cost-effective and scalable monitoring by utilizing existing fiber-optic communication infrastructure.

Additional advancements in railway diagnostics involve the integration of sensor technologies and real-time data analysis [5]. The “closed system” approach [6], which combines advanced sensors with centralized control architectures, improves responsiveness in infrastructure maintenance. Parallel efforts in intelligent traction power systems [7] aim to optimize power management and reduce capital investment through the optimal control of converters, energy storage devices, and renewable sources.

These solutions rely on synchronized voltage distribution measurements, achieved through low-power microcontrollers (e.g., Atmega128RFA1) and wireless transmission.

In the context of evolving security threats, the use of UAVs presents new vulnerabilities for critical infrastructure, including railway systems. The development of drone detection technologies is essential to counteract malicious UAV incursions. One study [8] introduced a long-range autonomous drone detection platform that integrates hardware and software components to ensure effective target localization. Its adaptive camera focal system and region-based neural network processing allow it to achieve a detection accuracy of 95.5% at distances up to 250 meters. The system is built on NVIDIA GeForce GTX 1080 and processes dual-stream media input with low latency and records intrusion events with contextual metadata.

Another promising approach [9] utilizes acoustic detection and deep learning algorithms for UAV identification. A hybrid dataset combining real recordings with synthetic audio generated by generative adversarial networks is proposed to overcome the scarcity of labeled drone audio data. Comparative testing of convolutional neural network (CNN), recurrent neural network (RNN), and convolutional recurrent neural network (CRNN) architectures demonstrates the hybrid dataset's effectiveness in recognizing both known and novel UAVs, confirming generative adversarial networks' potential to boost classification accuracy.

The growing ubiquity of drones in fields such as logistics, security, and engineering makes the need for automated detection systems more urgent than before. In [10], a YOLOv4-based detection model was trained on mixed drone and bird datasets. Performance evaluations using metrics such as mean average precision, precision, recall, F1-score, and FPS show that the system achieves high detection speed and accuracy, particularly when tested on DJI Phantom III and DJI Mavic Pro platforms.

These studies highlight the rapid advancement and integration of wide area monitoring systems and drone detection technologies in modern infrastructure protection. Leveraging AIoT platforms, cloud computing, and low-cost embedded systems enables the deployment of scalable, high-performance monitoring solutions. The inclusion of mechanical and environmental parameters improves system robustness, while deep learning and synthetic dataset generation significantly enhance drone detection accuracy. These innovations ensure timely fault identification, efficient maintenance planning, and resilience against aerial threats. In turn, they support the secure and sustainable operation of critical railway infrastructure.

3. ANALYSIS OF ACOUSTIC SIGNATURES OF AERIAL THREATS

The identification and classification of aerial threats based on their acoustic signatures is a critical component of developing effective real-time detection systems for critical infrastructure protection. This study gives particular attention to low-flying UAVs, which pose a substantial risk due to their ability to approach targets at low altitudes, which means they can often avoid conventional radar detection. These UAVs are typically mass-produced using cost-effective internal combustion engines, such as the MD-550 piston engine, which is widely used in improvised or commercially modified attack drones. Their flight generates a distinctive acoustic footprint - a combination of cyclic piston engine operation and rotor blade interactions with air - which is often perceived as a deep, repetitive buzzing or droning noise.

So that classification algorithms could be further refined, the dataset was expanded to include audio and video fragments of cruise missile flyovers. These threats exhibit a distinctly different acoustic profile, characterized by a more stable and continuous high-frequency noise spectrum caused by sustained jet propulsion. However, the number of verified recordings of cruise missiles was limited compared to the substantial volume of Shahed-type UAV data.

An essential step in this research involved the collection and preprocessing of audio samples from publicly available sources, including open social media channels. The gathered recordings were manually curated to extract clean acoustic segments; explosion sounds, speech, and other non-relevant noise artifacts were removed. Audio tracks were separated from video recordings and clipped to focus on flyover phases. Standard audio and video processing software tools (e.g., Audacity, FFmpeg) were

used during this stage. The processed dataset was then subjected to spectral analysis to isolate stable frequency components that consistently appear in recordings of specific aerial threats. An analysis was carried out in MATLAB using Audio Toolbox and DSP System Toolbox. This enabled high-resolution spectrogram generation and the statistical modeling of sound energy distributions over time and frequency. Acoustic signature extraction involved comparing multiple instances of UAV and missile flyovers, filtering out environmental noise, and identifying recurring spectral patterns. This enabled the derivation of distinctive acoustic fingerprints for each type of aerial threat. These fingerprints included dominant frequency bands, harmonic structures, and temporal dynamics of sound energy.

The resulting acoustic models account for variability in sound due to Doppler shifts, propagation through different environments, and recording device limitations. Despite these factors, consistent spectral density features were found across multiple samples, providing a reliable basis for signal classification. Figure 1 illustrates examples of raw acoustic signals and their corresponding frequency spectra for two distinct types of aerial threats, namely, a strike drone and a cruise missile. These signals illustrate the temporal waveform and spectral distribution at peak intensity. Figure 2 presents acoustic signatures extracted through spectral analysis and filtering. In the figure, panel (a) corresponds to a typical piston-engine UAV, and panel (b) illustrates the signature of a jet-powered cruise missile. The visual differences in harmonic structure and dominant frequency ranges between the two threat types form the basis for further classification and detection.

These findings form the basis for the acoustic recognition algorithms discussed in the following section and demonstrate the feasibility of threat-specific identification using low-cost, embedded microphone systems.

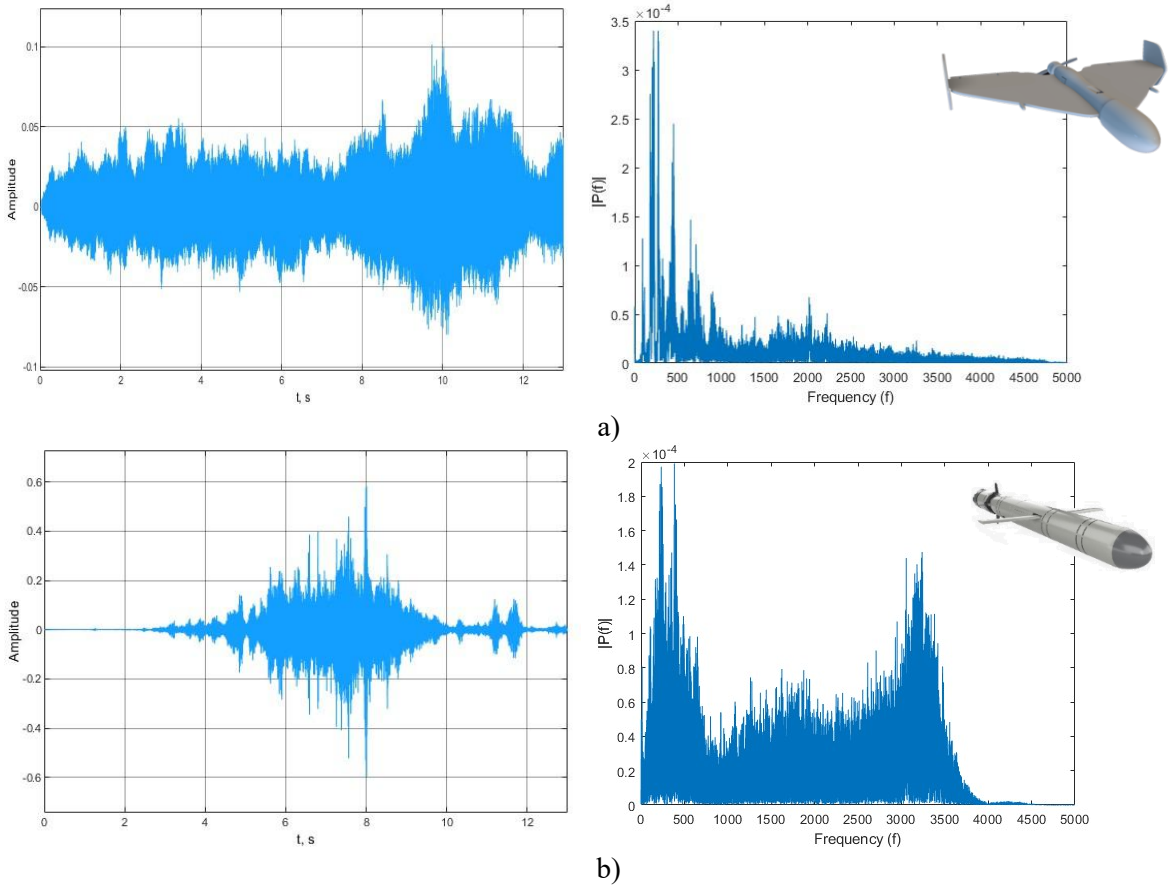
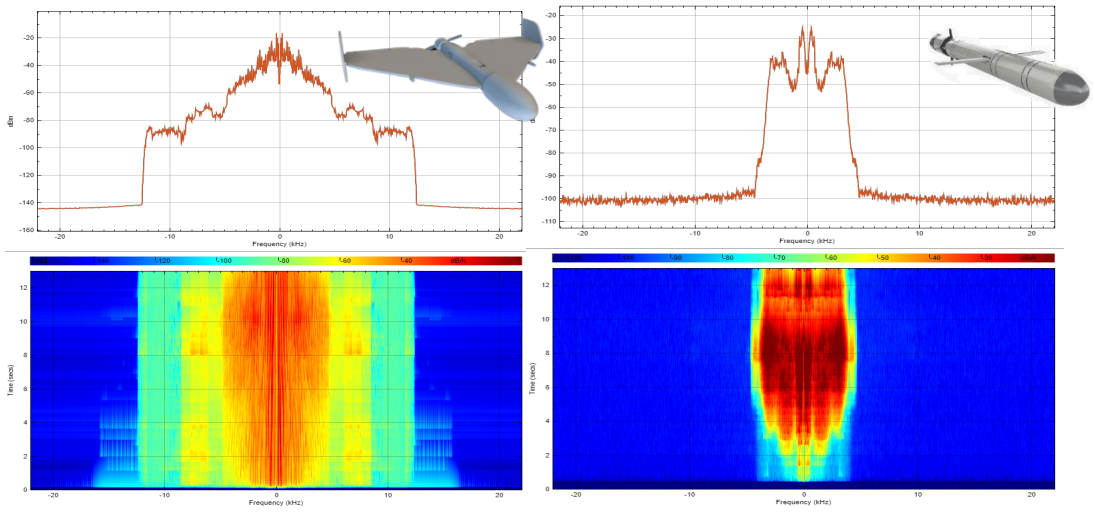


Fig. 1. Acoustic signal and its spectrum from (a) a strike drone and (b) a cruise missile



a) b)

Fig. 2. Acoustic signatures of a strike drone (a) and a cruise missile (b)

Table 1 summarizes the key spectral and temporal characteristics of acoustic emissions from UAVs and cruise missiles. This comparison highlights the distinct acoustic profiles of piston-engine drones, such as those based on MD-550 engines, versus jet-powered cruise missiles. Notable differences include dominant frequency ranges, harmonic structure, and sound modulation patterns, which are critical for effective threat classification. These parameters were derived from a statistical analysis of multiple recordings and serve as foundational features for Mel-frequency cepstral coefficient (MFCC) extraction and machine learning-based recognition.

Table 1

Comparative Acoustic Characteristics of Aerial Threats

Parameter	UAV (Shahed)	Cruise Missile (Kh-101, Kalibr)
Dominant Frequency Range	80-250 Hz	400-1000 Hz
Harmonic Components	Strong, periodic harmonics up to 2 kHz	Broad-spectrum, less structured
Modulation Pattern	Amplitude modulation due to blade rotation (3-8 Hz)	Relatively stable tone, minor amplitude variation
Spectral Density Peaks	Discrete peaks every ~80-120 Hz	Continuous elevated spectral band
Noise Type	Pulsed engine noise + blade harmonics	Sustained jet roar
Sound Pressure Variability (5s Window)	±6 dB fluctuation (due to modulation)	±2 dB fluctuation
Duration of Detectable Signature	5-15 seconds before flyover	3-8 seconds before flyover
Doppler Shift Presence	Moderate (variable path geometry)	Clear shift in tone with movement
Recording Sample Rate (for Capture)	≥16 kHz recommended	≥24 kHz preferred (for higher fidelity)
Effective MFCC Band Count	13-20	20-30

The dominant frequency range is the primary energy band in which the loudest components of the sound are concentrated, and it is a key indicator of the acoustic source type. Harmonic components

reflect the presence of periodic signals, often resulting from engine cycles and rotor blade interactions − these features are especially pronounced in piston-engine drones. Spectral density peaks represent sharp increases in acoustic energy at specific frequencies and are valuable for distinguishing between different types of aerial threats. Sound pressure variability describes the degree to which the amplitude of the acoustic signal fluctuates over time, capturing modulation effects that are characteristic of certain propulsion systems. Lastly, the MFCC band count indicates the number of Mel-frequency cepstral coefficients required to adequately represent the spectral features of the signal, which influence the effectiveness of subsequent machine learning classification.

4. AERIAL THREAT IDENTIFICATION METHODS

The classification of aerial threats based on acoustic data is a specialized case of environmental sound recognition, which is a well-established field in artificial intelligence. This problem involves identifying the type of sound event from an audio recording, often in the presence of background noise or environmental variability. In our case, the goal is to distinguish between UAVs, such as piston-engine drones, and other aerial threats, such as cruise missiles, based on their unique acoustic emissions captured in real-world conditions.

To achieve reliable recognition under such variability, the authors propose using deep learning models, which have proven to be effective for audio classification. These models automatically learn relevant patterns from feature representations such as spectrograms or cepstral coefficients. Widely used models include artificial neural networks (ANNs) and CNNs. These models offer a trade-off between complexity, training time, and deployment efficiency.

This research builds on the publicly available implementation of an audio classification system [11]. The authors acknowledge and thank the project's creator for open-source contributions that enabled adaptation to drone detection tasks. The referenced system offers a practical comparison of ANN, CNN1D, and CNN2D architectures on the UrbanSound8K dataset and demonstrates how such models can be trained and deployed with high usability. At the current stage of development, the ANN model has been selected, as it offers the highest classification probability for the targeted sound types and features a predictable, scalable architecture well-suited for embedded implementation.

Raw waveform signals must be transformed into meaningful compact representations so that audio data can be fed into machine learning models effectively. This study used MFCCs, a widely adopted feature in speech and sound recognition. MFCCs are designed to capture the perceptually relevant aspects of sound by mimicking the frequency resolution of the human auditory system.

One of the key advantages of MFCCs is their robustness when the durations of input sounds vary. Unlike raw waveforms or simple spectral features, MFCCs are computed over short, overlapping frames (e.g., 25 ms with 10 ms stride) and then averaged or pooled across time, enabling consistent feature dimensionality, even for audio signals of different lengths. This makes MFCCs well-suited for processing field-recorded UAV and missile audio segments, which may vary in duration from 2-10 seconds, depending on recording conditions.

The process of computing MFCCs involves the following steps:

1. Framing and windowing the signal into overlapping segments.
2. Computing the short-time Fourier transform for each frame.
3. Applying a Mel-scale filterbank to the power spectrum.
4. Taking the logarithm of the filtered energies.
5. Applying the discrete cosine transform to decorrelate the features.

The general formula for the *k-th* MFCC coefficient from a windowed signal frame is:

$$C_k = \sum_{n=1}^{N} \log(S_n) \cdot \cos\left[\frac{\pi k}{N}(n-0.5)\right], \quad k = 1,2,...,K \tag{1}$$

where S_n is the energy output of the *n-th* Mel filter, N is the number of Mel filters, and K is the number of retained cepstral coefficients (typically 13-20).

As a basic example, consider a one-second audio clip sampled at 16 kHz. Dividing it into 25-ms windows with a 10-ms overlap yields approximately 98 frames. The MFCC vector of 20 coefficients is computed for each frame, resulting in a 20×98 feature matrix. This matrix can then be averaged over time or passed as a time sequence into an ANN or CNN-based classifier.

The practical implementation involves a distributed acoustic detection system, composed of autonomous sensor units deployed in a grid along critical infrastructure zones. Each unit includes:

1. a microphone array for omnidirectional sound capture;
2. local preprocessing and MFCC extraction in real-time;
3. an embedded neural network model to classify the incoming audio frame;
4. wireless communication (e.g., ZigBee, LoRa) for forwarding alerts.

The system works as follows:

1. Acoustic data are continuously captured by each node;
2. MFCC features are computed and fed into a locally stored classifier;
3. When an aerial threat is detected, metadata (type, confidence, timestamp, possibly direction) is sent to a central hub;
4. Sensor data from multiple nodes can be fused to estimate UAV trajectory and speed.

Nodes must be low-power, self-sufficient, and designed for integration into a larger multi-modal alerting system. This architecture allows for the scalable and layered protection of railway and energy assets.

ANNs are the simplest architecture and are composed of fully connected layers. Input features, such as the averaged MFCC vector of fixed length (e.g., 20 coefficients), are passed through multiple dense layers with nonlinear activations (ReLU), culminating in a softmax output layer that yields probabilities for each threat class (e.g., UAV, cruise missile, background).

The dataset was augmented using several common audio data transformations to enhance model generalization and improve classification robustness in real-world environments:

a) Pitch shifting (±1-2 semitones) simulates variation in engine RPM;
b) Volume scaling (±3-5 dB) compensates for microphone gain or distance;
c) Background mixing adds ambient noise (wind, traffic, human speech);
d) Time stretching (±10%) models Doppler-like temporal distortions.

These augmentations ensure the neural network models are not overfitted to narrow recording conditions and can perform reliably in the presence of unpredictable background interference.

The average MFCC coefficients for a set of classified audio samples are presented in Table 2. Each audio class, including aerial threats (e.g., Shahed, missile) and environmental sounds (e.g., dog barking, engine idling), exhibits a distinct spectral profile. Despite some correlation between similar classes, all column vectors are unique, confirming that the MFCC feature set effectively captures discriminative characteristics. This supports the validity of MFCC-based representation for reliable audio classification.

Figure 3 illustrates the general architecture of the ANN used for audio-based classification. The input layer consists of 20 neurons, corresponding to the number of extracted MFCC features per audio segment. The output layer contains neurons equal to the number of target audio classes, enabling multi-class classification. The network includes multiple hidden layers, whose structures vary across configurations to balance classification performance and memory constraints.

The target hardware platform is based on an STM32 microcontroller with a total flash memory of 512 KiB. The number and size of hidden layers are optimized so that the complete network fits within about 450 KiB of program memory, allowing deployment on embedded systems without external memory.

The performance characteristics of several ANN configurations used for audio classification are summarized in Table 3. Each row in the table corresponds to a specific hidden layer architecture, and the number of neurons in each layer is listed in the first column. The second column shows the total model size in KiB, which is calculated based on the number of trainable parameters using 32-bit floating point precision. The third column presents the resulting validation accuracy achieved on the classification task. While the data generally show that increasing the number of neurons improves accuracy, this trend is not strictly monotonic due to the stochastic nature of training procedures,

including random weight initialization and mini-batch selection. This highlights the trade-off between classification performance and memory constraints, which is particularly important for deployment in embedded systems with limited flash storage, such as STM32 microcontrollers.

Table 2

Average MFCC Coefficients for Classified Audio Samples

MFCC No. (C_k)	Shahed	Missile	Dog Bark	Engine Idling	Jackhammer
1	-254.4	-163.6	-215.2	-125.8	-64.8
2	113.8	117.8	80.9	78.4	47.3
3	-28.5	-31.4	-24.5	-2.2	-2.4
4	15.0	22.6	-0.3	17.1	11.0
5	-8.3	-9.4	-9.9	0.4	-1.6
6	4.1	5.7	1.2	13.2	9.0
7	-8.5	-5.1	-7.2	0.7	-1.9
8	1.0	-0.5	3.2	8.2	4.5
9	-6.6	-3.1	-4.3	-1.4	-4.2
10	0.3	-1.2	4.2	6.6	4.5
11	-5.6	-3.0	-2.0	-1.2	-3.6
12	-0.1	-0.6	4.4	5.1	2.2
13	-5.4	-3.4	-1.8	-1.5	-3.3
14	-0.3	-0.7	4.2	3.7	2.0
15	-4.5	-2.8	-1.7	-2.1	-4.5
16	0.0	-0.9	3.4	3.2	1.5
17	-3.9	-2.0	-2.7	-3.1	-3.2
18	0.2	-1.0	2.5	1.9	2.7
19	-3.6	-1.8	-3.2	-2.7	-3.2
20	0.3	-0.7	2.3	2.3	2.2

The training dynamics of the selected ANN model, which show the evolution of classification accuracy over training epochs for both the training and validation datasets, are illustrated in Figure 4. The steady convergence and minimal gap between training and validation accuracy indicate that the model generalizes well without significant overfitting. The steady convergence and minimal gap between training and validation accuracy indicate that the model generalizes well without significant overfitting.

The training dataset was deliberately augmented by mixing samples with background noise, as well as varying volume and pitch, to further mitigate overfitting. This ensured sufficient variability and robustness in the training data. Training only on clean recordings leads to overfitting, in which case additional techniques, such as dropout, batch normalization, early stopping, or regularization, would be required. In this study, however, dataset augmentation proved sufficient to resolve the overfitting issue and maintain reliable model generalization.

The relationship between model size and classification accuracy for various ANN architectures is presented in Figure 5. As the figure shows, increasing the number of hidden layer neurons improves accuracy up to a certain point, beyond which the performance gains diminish. Within the practical range of ANN sizes between 300 and 500 KiB, the observed validation accuracy remains within 85-90%. This demonstrates that it is feasible to deploy reliable AI-based classification on microcontrollers such as the STM32 without exceeding available memory constraints.

5. HARDWARE AND SOFTWARE DEVELOPMENT OF AN EMBEDDED AI SOLUTION

The implementation of an embedded AI solution is centered around three core components: a micro-electro-mechanical systems (MEMS) microphone, an STM32 microcontroller, and a ZigBee wireless communication module. The MEMS microphone offers high sensitivity, compact size, and low power consumption, making it ideal for continuous acoustic monitoring in distributed sensor networks. The STM32 microcontroller serves as the central processing unit, selected for its balance of computational capability, low energy usage, and support for on-device neural network inference through STM32Cube.AI. The ZigBee module enables reliable, low-power wireless data transmission to higher-level systems or coordinators in mesh network configurations. Together, these components form a compact, cost-effective, and energy-efficient embedded solution tailored for field deployment in critical infrastructure environments.

Figure 6 is a schematic diagram of the hardware architecture designed for the embedded audio-based drone detection system.

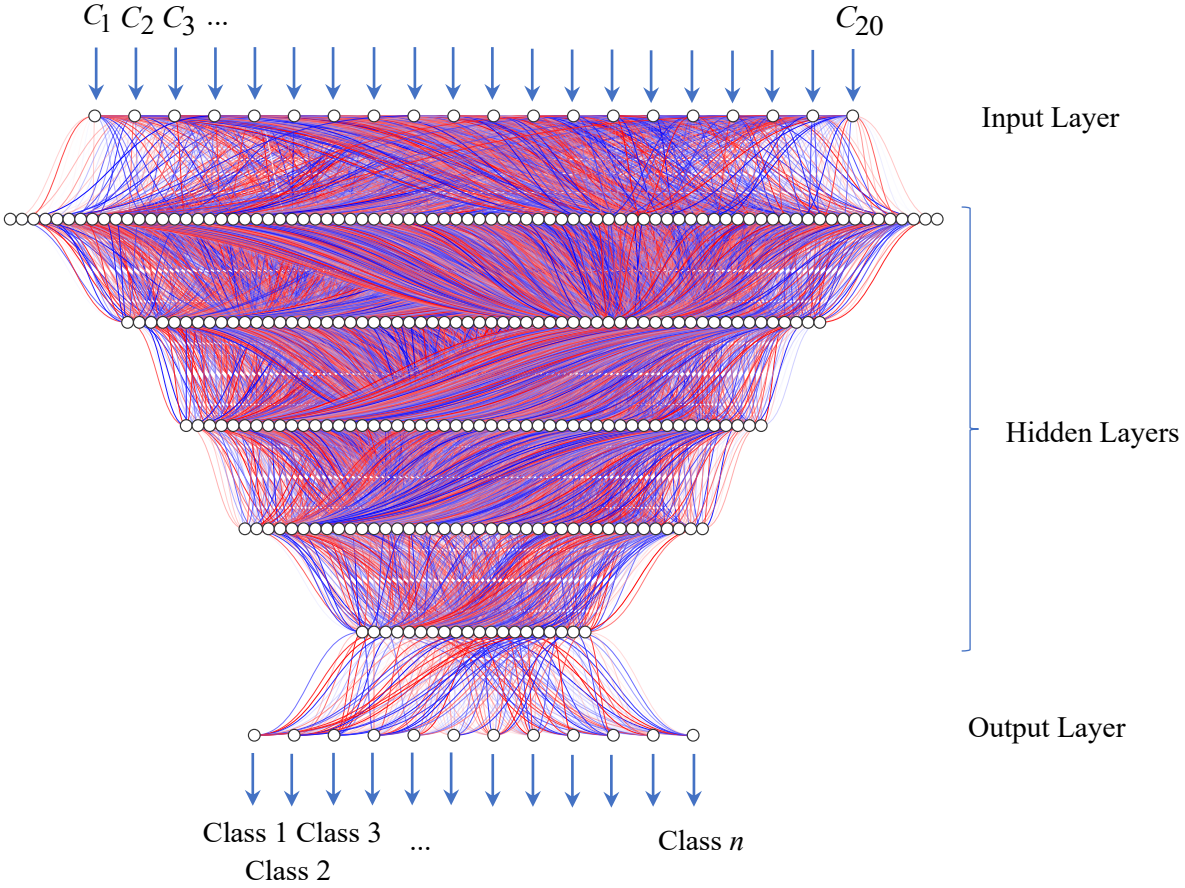


Fig. 3. General Architecture of the ANN Used for Audio-Based Classification

Fig. 7 illustrates the process of integrating the ANN into the MCU firmware using the STM32Cube code generation environment. The STM32 platform supports a modular approach through existing software libraries, enabling efficient development of embedded AI applications. For digital sound acquisition, the system leverages built-in support for MEMS microphones, utilizing PDM-to-PCM conversion and digital interfaces (I2S). Feature extraction is implemented using the CMSIS-DSP library, which provides optimized routines for computing MFCC coefficients in real time. The final classification stage is performed by the embedded ANN, which is converted and deployed using X-CUBE-AI, STMicroelectronics' toolkit for integrating trained neural networks into STM32 firmware. This software stack enables a complete on-device signal processing and inference pipeline with minimal resource overhead.

Table 3
Target ANN Model Size and Accuracy Depending on Hidden Layer Architecture

Architecture of ANN hidden layers	ANN Model Size	Validation Accuracy
1000-750-500-250-100-50	4.98 MiB	0.9046
512-256-128-64-32	737.5 KiB	0.8850
416-208-104-52-26	497.9 KiB	0.8931
400-200-100-50-25	462.7 KiB	0.8899
392-196-98-49-24	445.4 KiB	0.8959
384-192-96-48-24	428.7 KiB	0.8909
368-184-92-46-23	396.1 KiB	0.8822
352-176-88-44-22	364.8 KiB	0.8746
320-160-80-40-20	306.2 KiB	0.8866
256-128-64-32-16	204.9 KiB	0.8479
128-64-32-16	65.3 KiB	0.8026
32-16	16 KiB	0.6963

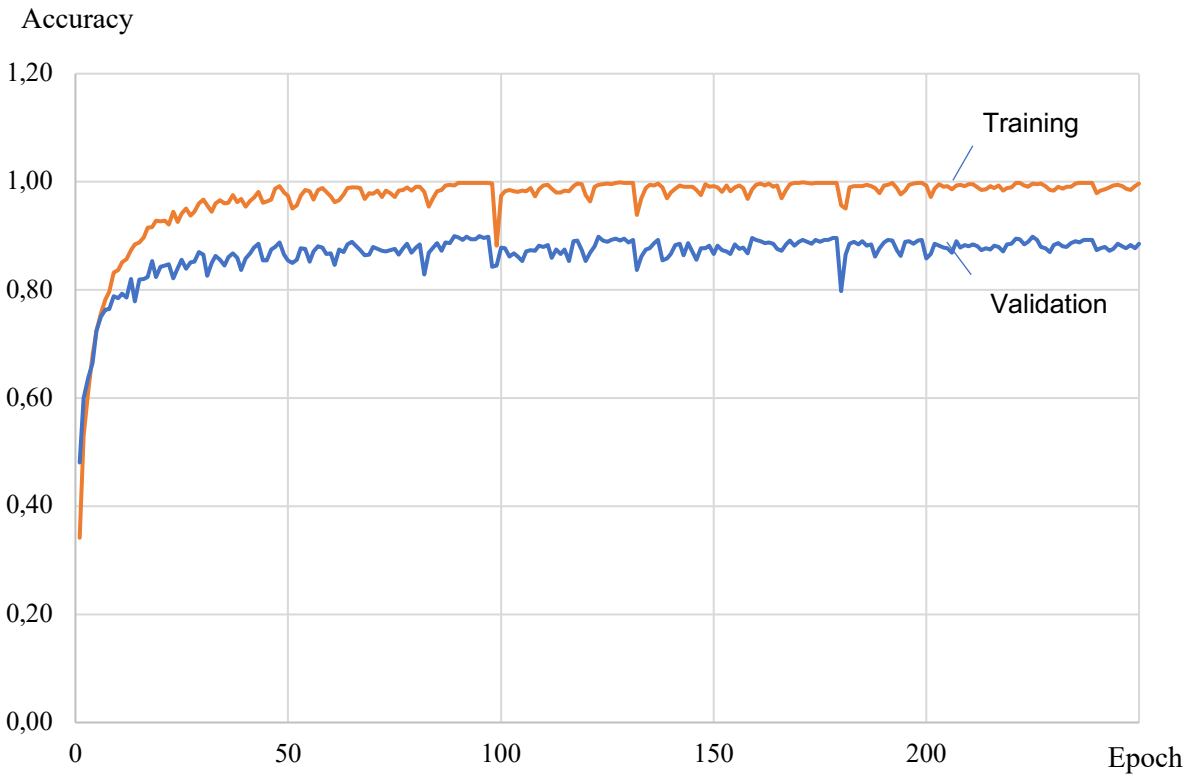


Fig. 4. Training and Validation Accuracy of the ANN

Based on the researched technologies, the authors propose a concept for railway infrastructure condition monitoring integrated with aerial threat detection (Fig. 8). The proposed system employs two wireless technologies: ZigBee - used for short-range communication between sensing units within a 75- to 100-meter radius - and LoRa - used for long-range transmission (5-15 km) of critical, summarized data such as anomaly events or confirmed aerial threats to a central control system.

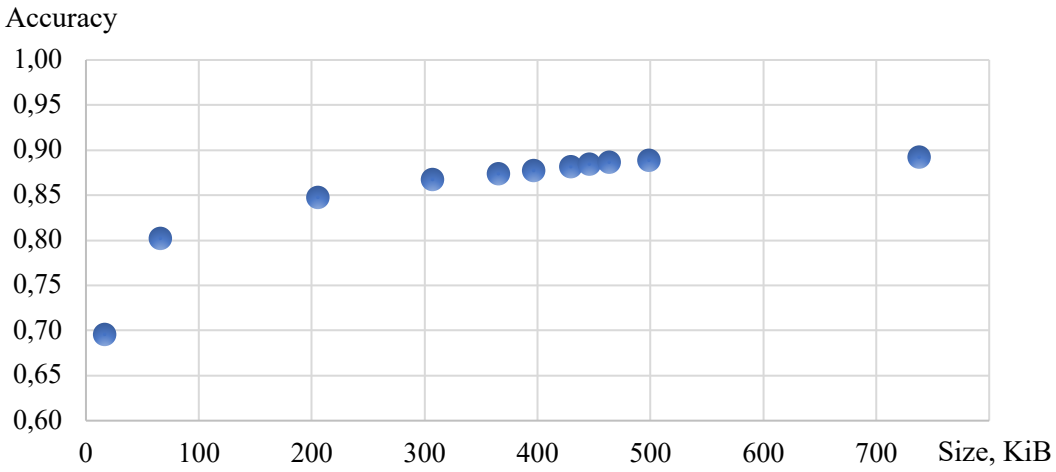


Fig. 5. Accuracy as a Function of Target ANN Model Size

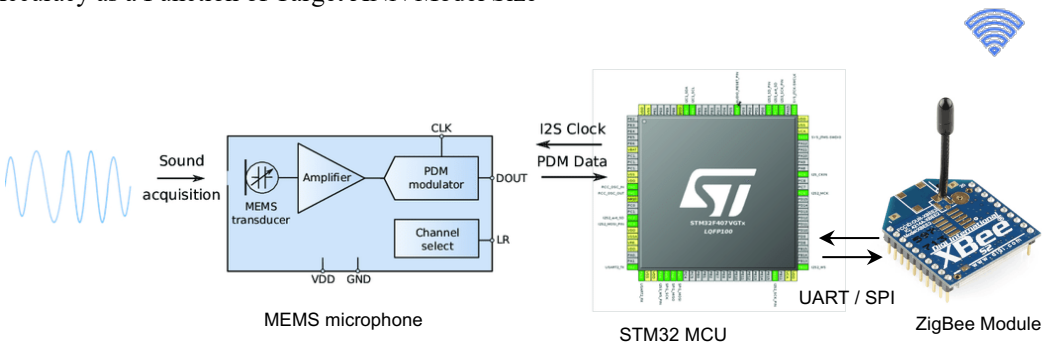


Fig. 6. Schematic Diagram of the Hardware Design for the Audio-Based Drone Detector

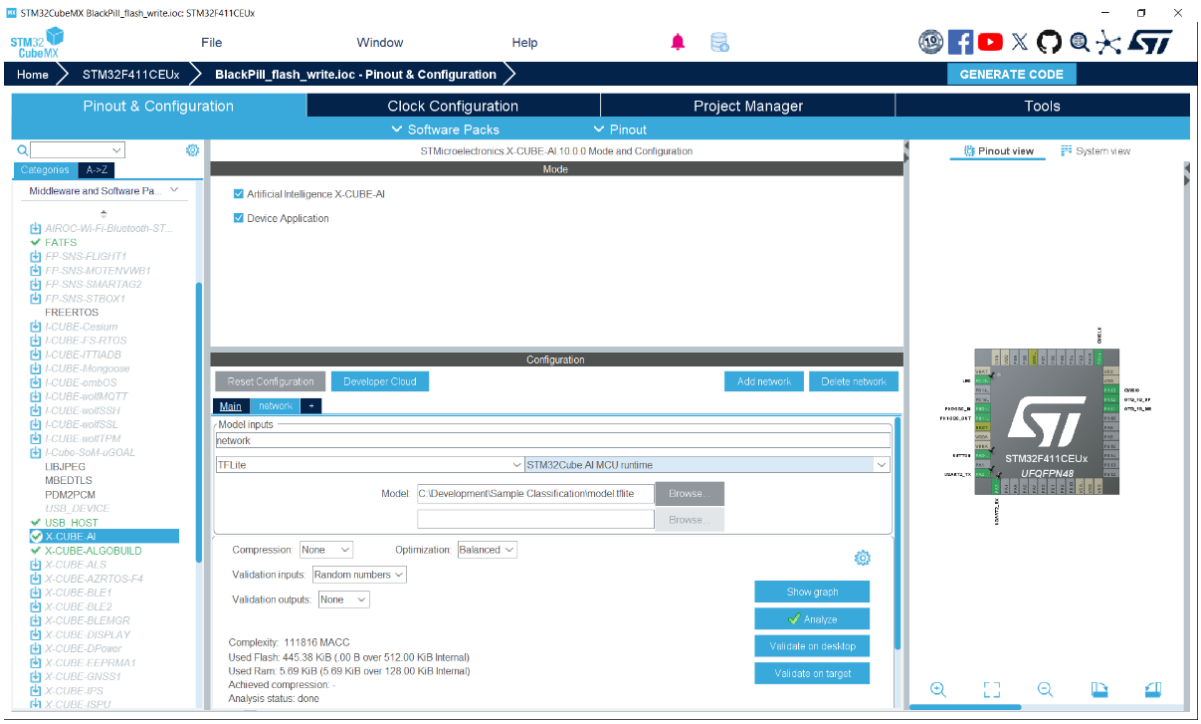


Fig. 7. Integration of the ANN into MCU Firmware Using the STM32Cube Code Initialization Tool

At the core of each sensing unit is a low-power STM32 microcontroller interfaced with a digital MEMS microphone, which can continuously monitor acoustic signals. Sound acquisition is initiated when the ambient noise exceeds a predefined amplitude threshold, reducing energy consumption and minimizing unnecessary processing. Once a signal is captured, it is processed locally: the microcontroller computes MFCCs using optimized CMSIS-DSP routines and then performs threat classification using an embedded ANN deployed via X-CUBE-AI.

The entire processing pipeline, from sound acquisition and feature extraction to inference, is executed on-device, without the need for external processing or connectivity. This embedded AI architecture ensures low latency, energy efficiency, and reliable performance in field conditions. The modular design allows the same platform to be extended to monitor other critical parameters of railway infrastructure, such as vibration, electrical noise, or structural anomalies, making it a scalable solution for next-generation intelligent railway systems.

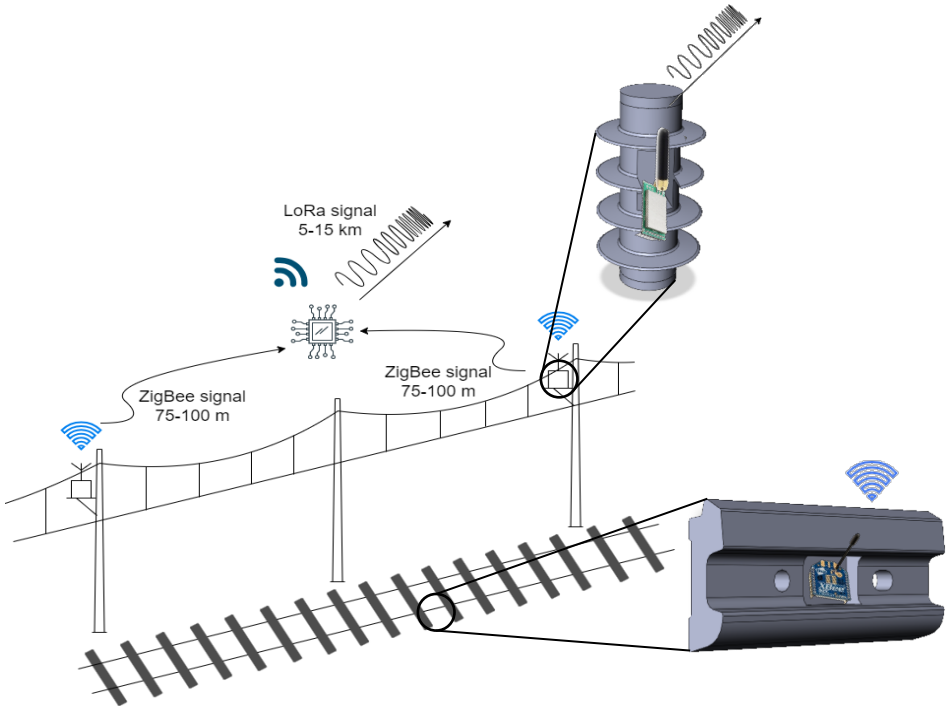


Fig. 8. Scheme of Railway Infrastructure Condition Conceptual Monitoring with Integrated Aerial Threat Detection

6. CONCLUSIONS

This study proposed a novel embedded AI-based system for real-time detection of aerial threats integrated into a broader railway infrastructure monitoring framework. The system design combines compact MEMS microphones, STM32 microcontrollers, and ZigBee/LoRa wireless communication modules to form a distributed network of acoustic sensors capable of localized sound capture and processing. Acoustic signals were analyzed on-device using Mel-frequency cepstral coefficients (MFCCs) as the primary feature set, followed by classification through lightweight ANNs deployed with the STM32 X-CUBE-AI toolkit.

This work advances the application of widely known methods (MFCC feature extraction and neural-network-based classification) through their adaptation, optimization, and validation in resource-constrained embedded environments. First, the study was based on real audio datasets, enabling practical feature classification and demonstrating the effectiveness of MFCCs in drone detection under realistic conditions. Second, a systematic investigation of embedded ANN architectures revealed the trade-off between model size and accuracy, showing that configurations in the 300- to 500-KiB range consistently achieve validation accuracy between 85% and 90%. This dependency between memory footprint and

classification performance provides valuable guidelines for embedded AI deployment. Third, the work presents the principal design of a low-cost, audio-based drone detection system built on commercially available, energy-efficient electronic components, indicating the feasibility of on-board neural network inference without reliance on external computing resources.

Beyond drone detection, the system's modular architecture enables it to be extended for broader railway condition-monitoring tasks, such as tracking electrical anomalies or mechanical wear in overhead line equipment. This study paves the way for scalable, intelligent, and autonomous sensing solutions for critical infrastructure protection by demonstrating how advanced signal processing and AI methods can be scientifically upgraded and practically adapted for STM32-class microcontrollers.

References

1. Surma, S. & Lukasik, J. et al. Contact network monitoring. *Electrification of Transport*. 2013. Vol. 5. P. 113-118.
2. Mizan, M. & Karwowski, K. & Karkosiński, D. et al. Monitoring odbieraków prądu w warunkach eksploatacyjnych na linii kolejowej. *Przegląd Elektrotechniczny*. 2013. R89. No. 12. P. 154-160. [In Polish: Monitoring current collectors in operating conditions on a railway line. *Electrotechnical Review*].
3. *Sicat CMS. Catenary monitoring system for overhead contact line systems*. Siemens Mobility GmbH. 2018. Available at: https://siemens.com/rail-electrification.
4. *Revolutionising Railway Decisions.* Sensonic GmbH. Available at: https://www.sensonic.com/.
5. Bondarenko, I. & Lukoševičius, V. & Keršys, R. et al. Innovative trends in railway condition monitoring. *Transportation Research Procedia*. 2024. Vol. 77. P. 10-17. DOI: 10.1016/j.trpro.2024.01.002.
6. Bondarenko, I. & Lukoševičius, V. & Neduzha, L. et al. Novel 'Closed'-system approach for monitoring the technical condition of railway tracks. *Sustainability.* 2024. Vol. 16(8). No. 3180. DOI: 10.3390/su16083180.
7. Bosyi, D. & Sablin, O. & Khomenko, I. et al. Intelligent technologies for efficient power supply in transport systems. *Transport Problems*. 2017. Vol. 12 (SE). P. 57-71. DOI: 10.20858/tp.2017.12.se.5.
8. Zhang, X. & Kusrini, K. et al. Autonomous long-range drone detection system for critical infrastructure safety. *Multimedia Tools and Applications*. 2021. Vol. 80. P. 23723-23743. DOI: 10.1007/s11042-020-10231-x.
9. Al-Emadi, S. & Al-Ali, A. et al. Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks. *Sensors.* 2021. Vol. 21(15). No. 4953. DOI: 10.3390/s21154953.
10. Singha, S. & Aydin, B. et al. Automated drone detection using YOLOv4. *Drones.* 2021. Vol. 5. No. 95. DOI: 10.3390/drones5030095.
11. Abishek, A.S. *Audio Classification Deep Learning*. Available at: https://github.com/abishek-as/Audio-Classification-Deep-Learning.
12. Yaran, W. & et al. Electrical interface of MEMS microphone introduction. *Infineon Knowledge Base Articles.* 2023. Jul 11. Available at: https://community.infineon.com/t5/Knowledge-Base-Articles/Electrical-interface-of-MEMS-microphone-introduction/ta-p/453658#.
13. Bosyi, D. & Sablin, O. & Potapchuk, I. Decentralized WAMS for railway overhead lines with audio-based drone detection. In: *2024 IEEE 5th KhPI Week on Advanced Technology (KhPIWeek).* Kharkiv: KhPI. 2024. P. 1-6. Available at: https://ieeexplore.ieee.org/document/10877981.

14. Kurhan, D. & Kovalchuk, V. & Markul, R. et al. Development of devices for long-term railway track condition monitoring: review of sensor varieties. *Acta Polytechnica Hungarica.* 2025. Vol. 22. No. 4. P. 65-82. DOI: 10.12700/APH.22.4.2025.4.5.

Received 20.04.2024; accepted in revised form 05.11.2025