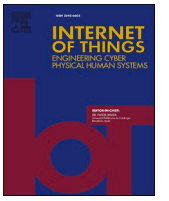
Internet of Things 36 (2026) 101864

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.sciencedirect.com/journal/internet-of-things



Scalable and low-power edge architecture with Wi-Fi HaLow and on-device spectrograms generation for flexible urban bioacoustics monitoring

Francisco A. Delgado-Rajó [a,*], Carlos M. Travieso-González [b], Ruyman Hernández-López [b]

[a] Institute for Technological Development and Innovation in Communications (IDeTIC), ULPGC Science and Technology Park, Edificio Polivalente II, 2ª planta. C/ Practicante Ignacio Rodríguez, s/n. 35017. Las Palmas de Gran Canaria, Spain

[b] Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC). Science and Technology Park, Edificio Polivalente II, 2ª planta. C/ Practicante Ignacio Rodríguez, s/n. 35017, Las Palmas de Gran Canaria, Spain

ARTICLE INFO

Keywords:
Birds' song recognition
Edge computing
Low-power wide-area networks
Biodiversity
Internet of things
Smart cities

ABSTRACT

Urban biodiversity monitoring in smart cities requires scalable and efficient computing architectures capable of handling real-time, distributed sensing tasks. This paper proposes a low-power edge computing and Internet of Things (IoT) framework that enables on-device acoustic detection and classification of bird species, serving as bioindicators of ecosystem health. The architecture leverages lightweight convolutional neural networks (CNNs) deployed on energy-efficient sensor nodes, significantly reducing communication overhead by transmitting only detection events and compact spectrogram data. A key contribution is the automatic generation of Mel-spectrograms at the edge, which supports the continuous creation of training datasets and iterative neural network refinement without manual preprocessing. The proposed system incorporates dual Wi-Fi and Wi-Fi HaLow connectivity, providing adaptable long-range, low-power communication for heterogeneous urban environments. Field experiments validate the framework's scalability and effectiveness, demonstrating robust detection of both native and invasive species. By combining distributed intelligence, resource-aware computation, and flexible networking, the system offers a practical edge–IoT solution for large-scale, real-time environmental monitoring in smart city contexts.

1. Introduction

In recent years, biodiversity monitoring has become an increasingly important discipline for assessing environmental change and ecosystem health. Within the context of smart cities, biodiversity indicators provide valuable insights into the effects of urban dynamics such as the creation of green areas, pedestrian zones, and the consequences of climate change [1,2]. These indicators help evaluate the environmental impacts of urban planning and decision-making processes.

One effective approach for monitoring urban biodiversity is the observation of bird species, both in terms of their seasonal presence and spatial distribution across metropolitan and suburban environments [3,4]. Birds serve as reliable bioindicators due to their

* Corresponding author.
E-mail address: paco.rajo@ulpgc.es (F.A. Delgado-Rajó).

https://doi.org/10.1016/j.iot.2025.101864

Available online 23 December 2025

sensitivity to changes in environmental conditions and habitat quality. For example, studies have demonstrated that bird diversity is strongly influenced by vegetation density and landscape configuration, which can be effectively assessed using indices such as the Normalized Difference Vegetation Index (NDVI) [5].

The integration of advanced technologies, including artificial intelligence (AI), acoustic monitoring systems, and citizen science, is transforming biodiversity monitoring [6,7], (Boussard et al., 2021). Smart cities are increasingly incorporating these digital tools into adaptive management frameworks that leverage satellite imagery, Internet of Things (IoT) devices, and mobile applications for real-time species tracking. These approaches not only improve the quality and granularity of ecological data but also enhance community engagement in urban conservation initiatives.

Currently, AI- and neural network–based techniques are widely employed to detect bird species through vocalization analysis. Most existing approaches rely on collections of audio segments obtained from public databases or through field recordings, while others focus on real-time audio processing [8,9]. In this study, audio segments are generated after the detection phase, serving as training material for model refinement tailored to each specific site. However, real-time applications often require continuous audio streaming, which leads to significant bandwidth consumption. To address this limitation, edge computing has emerged as an effective solution. By processing data locally on distributed edge devices, it is possible to significantly reduce network loads while enabling real-time bird detection and species classification without transmitting raw audio [1,9,10]. This is particularly advantageous for urban biodiversity monitoring, where multiple low-cost, low-power devices can be deployed to perform on-device inference in a distributed manner.

Nevertheless, existing CNN-based approaches often demand high computational resources and power due to their reliance on large convolutional layers and extensive matrix operations. These characteristics limit their scalability and make them unsuitable for continuous real-time operation on low-power edge devices, where memory, processing capacity, and energy availability are constrained. Consequently, achieving both scalability and low-power performance requires alternative architectural and deployment strategies, as proposed in this work.

This study presents an edge computing–based architecture for the monitoring of urban bird populations. Rather than transmitting continuous audio streams, the system issues detection alerts only when a target bird species is identified. This design significantly reduces communication overhead and enables the use of low-power wide-area network (LPWAN) protocols such as LoRa, LTE, or NB-IoT, which are well suited for battery-powered sensor networks [11,12]. Consequently, the system achieves both scalability and energy efficiency while supporting continuous and distributed monitoring across urban environments.

In this context, we incorporate Wi-Fi HaLow (IEEE 802.11ah), which provides extended range and lower power consumption compared to conventional Wi-Fi. Moreover, many urban parks and green spaces are already equipped with Wi-Fi infrastructure, facilitating direct integration with IP networks and enabling the deployment of additional functionalities. Our architecture dynamically employs both conventional Wi-Fi and Wi-Fi HaLow technologies, depending on the geographic characteristics of the deployment area. Both transmission protocols support TCP/IP connectivity, which is essential for the reliable transfer of small but critical data files.

Typically, datasets used for training neural networks are composed of audio segments containing only a small percentage of bird vocalizations [13,14]. This requires segmentation of audio files, often performed manually. During network training, the input is the segmented audio from which features (e.g., spectrograms) are extracted to feed the classifier [15].

One of the key contributions of the proposed system is its ability to automatically extract relevant sound segments for training new convolutional neural networks (CNNs) and transmit them to a centralized database for future use. This functionality addresses a major limitation of existing bird song datasets, which often require manual segmentation at the precise moment of vocalization. In contrast, our system not only detects bird species in real time but also autonomously generates Mel-spectrograms, thereby streamlining the creation of datasets for subsequent AI training. By producing these representations directly, future CNN models can be trained without needing a separate spectrogram generation step [16], significantly reducing preprocessing demands and overall training time. This automated feedback loop enables the system to continually learn and adapt, improving accuracy as more data is collected and new vocalizations are identified in the urban environment. Furthermore, this approach aligns with recent developments in continuous learning systems [17], which enable long-term adaptive training. Such systems, including ours, progressively refine their performance through successive iterations, complementing the system's low-energy design achieved through on-node processing and selective data transmission.

An especially important use case is the monitoring of invasive species such as the rose-ringed parakeet (*Psittacula krameri*), increasingly prevalent in European cities. Due to their loud and distinctive calls, these birds are particularly suitable for automated detection via trained neural networks. Early detection using edge AI is crucial for enabling a rapid response by local authorities, helping mitigate the negative impacts of these species on native biodiversity, especially in areas where ecological oversight is limited [18]. Additionally, monitoring other species supports the assessment of invasive–native interactions.

2. Related work

Recent studies have demonstrated the potential of deep learning for acoustic monitoring [19]. BirdNET, developed by Kahl et al. [20], employs convolutional neural networks (CNNs) to identify species from audio recordings, highlighting the effectiveness of spectrogram-based analysis. Other systems use visual detection through cameras and neural networks; however, acoustic identification is generally more efficient for species recognition and requires less bandwidth, as it avoids video transmission [21]. Early approaches relied on deterministic techniques such as pattern matching or hidden Markov models [22–24]. With advances in artificial intelligence, research has shifted toward CNN-based models that enable more accurate and automated recognition of species-specific vocalizations.

2.1. CNN networks

Despite these advances, most solutions have been designed for high-performance computing platforms, which limits their applicability on low-power edge devices. Goitia-Urdiain et al. [25] observed that the accuracy of automatic bird song recognition varies depending on the software used, highlighting the importance of robust methodologies and recordings across diverse environments to minimize misinterpretations of bird abundance and vocal activity.

The transformation of acoustic signals into deep learning–compatible features constitutes a critical preprocessing stage in automated bird identification systems. Mel-spectrograms have emerged as the predominant time–frequency representation, widely applied across numerous classification frameworks . Spectrogram processing through CNNs offers notable advantages for embedded and distributed systems, as it allows the use of compressed audio inputs while maintaining high classification accuracy. Chandu et al. [26] emphasized the potential of lightweight architectures for field deployment, enabling portability and energy efficiency. Nevertheless, dataset construction remains a challenge, as segmentation of audio fragments is often labor-intensive.

Several studies have combined IoT principles with bioacoustic monitoring. For example, Arowolo et al. [27] integrated embedded devices with communication techniques that bring computation closer to the node, reducing network load and system latency. This approach facilitates the development of portable, low-cost, and low-power systems, enabling effective real-time monitoring. More recently, Segura-Garcia et al. [28] presented a 5G AI–IoT system for bird species monitoring using CNN architectures trained on spectrograms. Their study compared EfficientNet and MobileNet models pretrained on ImageNet with lightweight CNNs, achieving up to 75 % accuracy. Although suitable for single-board computers (SBCs) and microcontroller units (MCUs), this system depends on mobile network coverage, which is not always available in remote or forested areas. Furthermore, it requires clean and well-segmented audio-to-image conversion, which increases preprocessing complexity. By contrast, our system automates Mel-spectrogram generation, thereby streamlining preprocessing and enhancing real-time applicability.

While dataset construction remains a central challenge, we also contextualized our proposed five-layer CNN against recent lightweight convolutional architectures. Table 1 summarizes model complexity and published performance for MobileNetV2 and EfficientNet-B0, compared to our design.

Our CNN includes only 0.06 million learnable parameters, whereas MobileNetV2 and EfficientNet-B0 contain 3.4 M and 5.3 M parameters, respectively—approximately 50–90 times larger. Although these architectures achieve higher ImageNet accuracy (71.8 % and 77.1 %), their computational cost (300–390 M FLOPs) makes them less practical for small-scale, domain-specific datasets. In contrast, our compact model achieves efficient training and inference with minimal risk of overfitting, making it suitable for resource-constrained or real-time scenarios.

2.2. Edge computing

Shi et al. [31] first conceptualized edge computing as a strategy to reduce latency and bandwidth consumption by executing inference directly on sensor nodes rather than transmitting raw data to the cloud. Recent IoT-based agricultural systems have also leveraged edge computing and context-aware sensing to optimize environmental management. For example, Khan et al. [32] proposed an IoT-assisted context-aware fertilizer recommendation framework that adapts nutrient supply to soil and crop conditions. Related works include the intelligent optimization of reference evapotranspiration (ETo) for precision irrigation and context-aware evapotranspiration (ETs) modeling for saline soil reclamation [32], which demonstrate how IoT and lightweight AI can enhance sustainability and decision making in field deployments.

Our work extends this paradigm to bioacoustics monitoring in smart cities by embedding lightweight CNNs within resource-constrained IoT nodes. While **model complexity** and **data privacy** are important considerations in ecological monitoring, this study focuses on the **implementation of a bioacoustics monitoring platform**. Future work will address these aspects through the integration of more sophisticated models and the adoption of privacy-preserving methods, such as federated learning, to ensure the protection of sensitive data and improve overall system efficiency.

The continuous evolution of embedded platforms in terms of cost, processing power, and connectivity has facilitated their use in ecological monitoring. For instance, Prinz et al. [33] developed a low-cost nest monitoring camera using a Raspberry Pi, infrared camera, and motion sensors, automatically uploading videos via Wi-Fi. Similarly, [34] implemented an RFID-enabled bird feeder with Raspberry Pi Zero W and PIT-tagged birds, transmitting visitation data (bird ID, date, time), with the option to integrate audio or video sensors. These examples demonstrate how AI-enabled devices can support distributed, field-based ecological monitoring.

2.3. Communication networks

In edge-processing systems, only detection alarms are transmitted, eliminating the need for continuous real-time audio streaming.

Table 1
Model complexity and published performance for MobileNetV2 and EfficientNet-B0, compared to our design.

Model	Params (M)	FLOPs (M)	Layers	Kernel Size	Activation	Top-1 Accuracy (ImageNet, %)	Reference
Five-layer CNN (ours)	0.06	~20–30	5 Conv + 2 FC	3 × 3	ReLU	96.69	This work
MobileNetV2	3.4	300	53	3 × 3 depthwise	ReLU6	71.8	[29]
EfficientNet-B0	5.3	390	82	3 × 3 / 5 × 5	Swish	77.1	[30]

Communication technologies such as NB-IoT, LTE-M (Cat-M1), and LoRa are particularly suitable for this purpose [28,35]. These protocols enable network scalability thanks to their low coupling and stateless communication models, which also improve fault tolerance. They further support telemetry protocols such as MQTT, facilitating the rapid integration or removal of sensor nodes.

Although low latency is not a strict requirement, cellular-based solutions depend on mobile network coverage. In contrast, LoRa offers long-range transmission, making it suitable for obstructed, suburban, or rural environments [36,37]. Adaptive protocols and hierarchical architectures can extend LoRa coverage through multi-hop communication [11,38]. This is possible due to its physical layer's use of Chirp Spread Spectrum (CSS) modulation [39].

In edge-processing systems, only detection alarms are transmitted, eliminating the need for continuous real-time audio streaming. Communication technologies such as NB-IoT, LTE-M (Cat-M1), and LoRa are particularly suitable for this purpose [28,35]. These protocols enable network scalability thanks to their low coupling and stateless communication models, which also improve fault tolerance. They further support telemetry protocols such as MQTT, facilitating the rapid integration or removal of sensor nodes.

Although low latency is not a strict requirement, cellular-based solutions depend on mobile network coverage. In contrast, LoRa offers long-range transmission, making it suitable for obstructed, suburban, or rural environments [36,37]. Adaptive protocols and hierarchical architectures can extend LoRa coverage through multi-hop communication [11,38]. This is possible due to its physical layer's use of Chirp Spread Spectrum (CSS) modulation [39].

Recently, Wi-Fi HaLow (IEEE 802.11ah) has emerged as a promising option for urban biodiversity monitoring, providing long-range connectivity (up to 1–2 km in dense cities), native IP compatibility, and lower power consumption than conventional Wi-Fi [40,41]. Unlike LoRa or NB-IoT, HaLow supports higher data rates (15–20 Mbps), allowing more complex payloads such as spectrogram transmissions or firmware updates without compromising energy efficiency. These features make it particularly suitable for smart city deployments where nodes must operate autonomously in heterogeneous environments.

In this work, we explore the use of Wi-Fi HaLow in scenarios where conventional Wi-Fi coverage is inconsistent. By leveraging its extended range and scalability, nodes can sustain efficient cloud communication while preserving low energy consumption. This reduces infrastructure requirements (e.g., access point density), thereby lowering deployment and maintenance costs for large-scale biodiversity networks.

Recently, Wi-Fi HaLow (IEEE 802.11ah) has emerged as a promising option for urban biodiversity monitoring, providing long-range connectivity (up to 1–2 km in dense cities), native IP compatibility, and lower power consumption than conventional Wi-Fi [40,41]. Unlike LoRa or NB-IoT, HaLow supports higher data rates (15–20 Mbps), allowing more complex payloads such as spectrogram transmissions or firmware updates without compromising energy efficiency. These features make it particularly suitable for smart city deployments where nodes must operate autonomously in heterogeneous environments.

In this work, we explore the use of Wi-Fi HaLow in scenarios where conventional Wi-Fi coverage is inconsistent. By leveraging its extended range and scalability, nodes can sustain efficient cloud communication while preserving low energy consumption. This reduces infrastructure requirements (e.g., access point density), thereby lowering deployment and maintenance costs for large-scale biodiversity networks.

2.4. Summary

Although enabling technologies such as CNN-based acoustic detection, low-power edge computing, and long-range wireless networking are well established, several recent studies have proposed integrated 5G–AI–IoT frameworks for environmental or bioacoustic monitoring e.g. [27,28].

However, few of these frameworks specifically target urban biodiversity monitoring or emphasize modular integration with citizen science infrastructures.

Our proposed architecture addresses this gap by combining edge AI, Wi-Fi/Wi-Fi HaLow connectivity, and modular integration with citizen science infrastructures to deliver a scalable, energy-efficient, and real-time bird detection system for smart cities.

The main contributions of this work are:

- Development of a low-cost, low-power IoT network of nodes that employs edge computing for real-time bird species detection.
- Flexible network connectivity, with nodes transmitting only alerts and one-second spectrograms of detected calls when Wi-Fi is available.
- Creation of a spectrogram database that supports the training of CNNs with larger and more diverse datasets.
- Reduction in training time by eliminating manual segmentation and using spectrograms directly as CNN inputs.

3. Materials and methods

The primary objective of this work is the deployment of nodes using Raspberry Pi devices as their core component. The Raspberry Pi has been selected due to its processing capabilities, which enable it to deploy a lightweight CNN network, and compatibility with a variety of microphones and external communication modules, such as the Wi-Fi HaLow transceiver.

The overarching goal is to monitor urban parks, which typically offer straightforward access to Wi-Fi access points, as well as more remote areas where alternative communication technologies, specifically Wide-area Low Power Wireless Networks (LPWANs) are required. A typical example is shown in Fig. 1.

3.1. System architecture

In scenarios where IEEE 802.11 (Wi-Fi) connectivity is available, direct communication with cloud platforms can be established, enabling real-time data visualization and access via TCP/IP protocols. This connectivity also allows for the transmission of small data files to cloud storage or a remote server. In more remote areas or larger areas, the use of HaLow Wi-Fi access points is proposed, which allow the connection of up to 100 nodes each.

The same node can employ both types of communication interchangeably, owing to the modularity of the system architecture, which facilitates network deployment. In the case of Wi-Fi HaLow, it offers four bandwidth modes (1, 2, 4, and 8 MHz), supporting transmission speeds of up to 32.5 Mbps under optimal conditions. The selection of the appropriate bandwidth mode is automatically managed by the Wi-Fi HaLow device, based on link quality, distance, and environmental conditions. Therefore, performance may vary depending on distance and the specific characteristics of the urban environment. Mode 1 corresponds to the lowest transmission speed, offering the greatest range.

The proposed network architecture supports two types of communication between the end nodes and the cloud. As a data visualization platform, Thingspeak [42] was selected due to its compatibility with data transmitted over the Wi-Fi network. It enables the visualization of node locations by assigning a dedicated channel to each node, with individual fields corresponding to each bird species detected.

The general architecture of the system is presented in Fig. 2. In each case, HaLow Wi-Fi technology is or is not used depending on the proximity of a conventional 802.11 access point. In the case of locations with Wi-Fi coverage (public parks, campus), it is not necessary, and the operation is similar for both systems.

Table 1 presents a comparison between the two communication technologies.

3.2. Monitoring node design

The central component of the monitoring nodes is a Raspberry Pi 3+, which enables the implementation of the Detection and Classification System, communication functionalities, alert processing based on severity levels, and the transmission of data packets using either of the two communication technologies. In the case of 802.11ah technology, a station communications node is connected via Ethernet using a bridge configuration that links the conventional network with the HaLow network.

With this implementation, the network connection is transparent to the embedded system used if it supports TCP/IP. The entire system deployed on the raspberry is developed in Matlab's Simulink [42]. This includes the capture of the sound by means of a USB microphone connected to the device, the extraction of MEL spectrograms and the convolutional neural network (CNN), in addition to the logic that publishes the alerts on the cloud platform or sends the spectrogram via a UDP packet to the database deployed locally on the server. The outputs of the neural network are the detection probabilities of each bird species normalized to 1. A threshold value is chosen from which the presence detection is published and the last spectrogram corresponding to one second of edge is sent. The block diagram of the system embedded in the Raspberry is shown in Fig. 3.

The system input is a microphone, followed by a pre-emphasis block, which precedes the feature extraction module discussed later. Each time a presence alert is triggered, all outputs from the neural network are published to Thingspeak, with each value assigned to a distinct field within the channel. However, only the spectrogram corresponding to the detected class (specifically, the last second of audio) is transmitted. To achieve this, Simulink implements a buffer block that stores the spectrogram of the most recent second of audio by reshaping the 98×50 matrix into a 1×4900 vector, which is transmitted if it corresponds to a detection event.

To assess the real-time performance of the Simulink-based processing pipeline on low-power hardware, the model was deployed on a Raspberry Pi 3 Model B (1.2 GHz quad-core ARM Cortex-A53, 1GB RAM) using the Simulink Support Package for Raspberry Pi Hardware.

Profiling Results:

- Average CPU Utilization: 62 %

Fig. 1. Typical scenery.

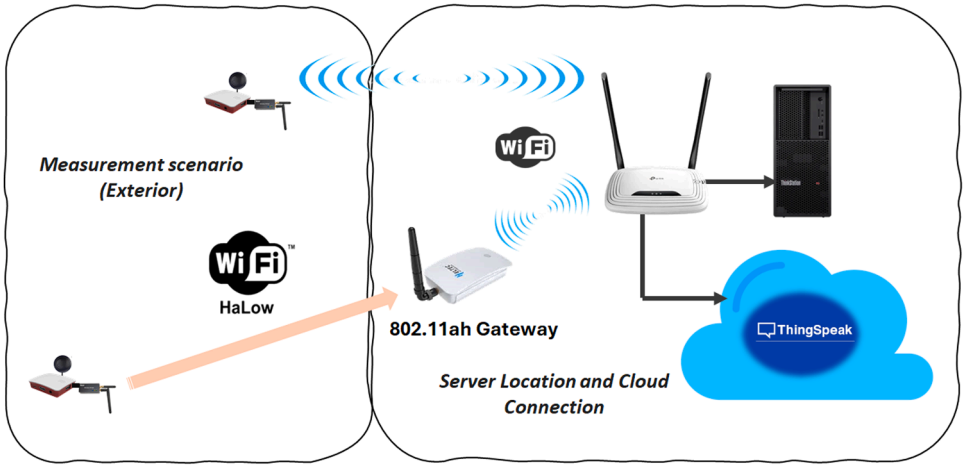


Fig. 2. System architecture.

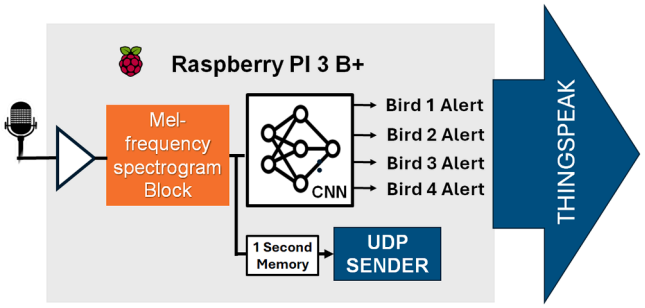


Fig. 3. Block diagram of the embedded system.

- Peak CPU Utilization: 78 % during concurrent data acquisition and FFT computation
- Maximum Observed Latency: 9.8 ms (within the 10 ms cycle period)

The node maintained stable real-time operation for over 90 min of continuous execution without missed deadlines or buffer overflows. These results confirm that the proposed Simulink-based design is computationally feasible for deployment on low-power embedded platforms such as the Raspberry Pi 3 Model B

The UDP packet containing this spectrogram consists of the 98×50 matrix of double-precision points reshaped into a 1×4900 vector solely for transmission purposes, preserving all frequency-time information. This allows the spectrogram to be sent efficiently over the network while maintaining the integrity of the data. Upon reception, the original 2D structure can be reconstructed for visualization or further processing. The flowchart is presented in Fig. 4.

One-second spectrograms corresponding to detection events are sent via UDP to the local server, where they are stored by species. This real-time transmission offers two main advantages: first, it enables the continuous and automatic creation of a growing, species-organized database for periodic retraining of the CNN models deployed on the edge nodes, allowing the system to progressively adapt to changes in the acoustic environment. Second, by transmitting spectrograms directly rather than raw audio, the system avoids the need for computationally expensive and error-prone audio segmentation to isolate bird vocalizations, significantly reducing pre-processing complexity and the computational load on resource-constrained devices. As more data accumulates, new versions of the model with improved accuracy and expanded detection capabilities can be generated. These model firmware updates could be remotely deployed to the nodes through Wi-Fi HaLow's network capabilities, ensuring that the system refines itself over time and adapts to the changing dynamics of the urban environment.

3.3. CNN network

The neural network implemented on the Raspberry Pi is a five-layer convolutional neural network (CNN) composed of simple convolutional layers responsible for filtering the input data, each followed by an activation function. This network is integrated into a Simulink model, which is deployed on the device using a .mat file. These types of networks are primarily used for image detection; in this case, the input is the spectrogram of a one-second audio segment corresponding to a specific bird species.

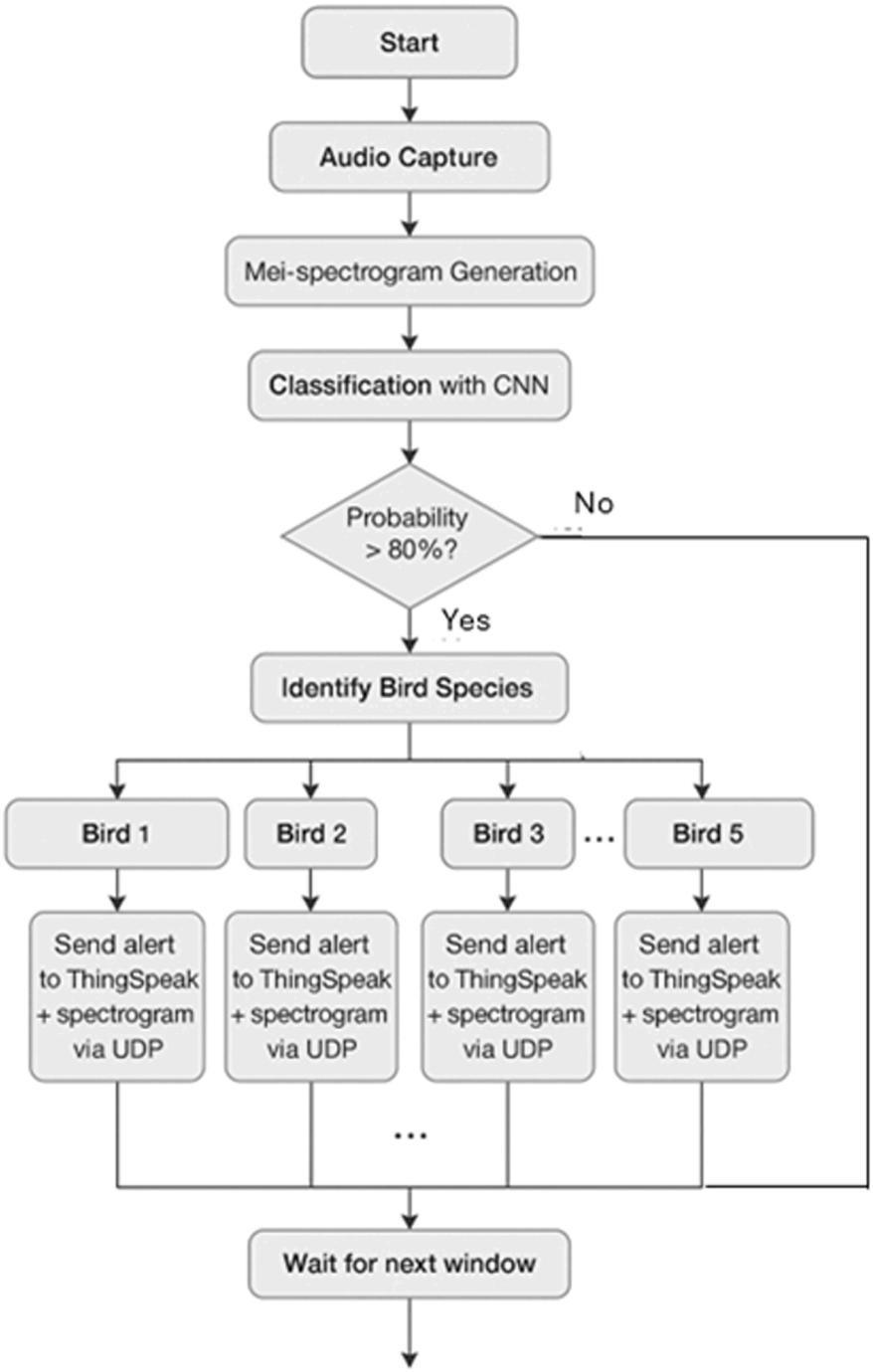


Fig. 4. System flowchart.

At the output of the network, a fully connected layer is responsible for generating the final predictions, based on the probability of occurrence of a given species according to its song.

The subsequent convolutional layers apply additional sets of filters, typically more complex than those in the initial layer. These filters are designed to detect higher-level features, such as textures and patterns, composed of the low-level features extracted earlier. Finally, the fully connected layer abstracts the features extracted by the convolutional layers and uses them to make the final prediction.

This network has been trained using data from various bird species that are commonly found in the parks where the field tests will be conducted. Therefore, the initial step in the training process involved standardizing the audio files from a public database to a single

format (specifically, mono-channel, WAV files of 1 s duration). This preprocessing step is no longer necessary in our system once spectrograms captured in situ are used as the training dataset for future networks. This implies savings in the segmentation process and in the reduction of stages in the processing of the training. Fig. 5 shows the previous process that is carried out for the training of the neural network from the audio database. This would be the usual one in almost all jobs. In this case it would only be done the first time.

The species selected for this study are those most frequently observed in urban parks in Las Palmas de Gran Canaria: the Heineken Eurasian Blackcap (*Sylvia atricapilla heineken*), the Canary Islands Chiffchaff (*Phylloscopus canariensis*), the Common Blackbird (*Turdus merula*), the Spanish Sparrow (*Passer hispaniolensis)* and, finally, the Rose-ringed Parakeet (*Psittacula krameri*). The latter is an invasive species in many European cities and is known to negatively impact native species by interfering with their nesting or feeding habits, as previously discussed.

One-second spectrograms corresponding to detection events are sent via UDP to the local server, where they are stored by species. This continuously growing database allows for the periodic retraining of the CNN models implemented on the nodes. As more data accumulates, new versions of the model with improved accuracy and expanded detection capabilities can be generated. These model firmware updates could be remotely deployed to the nodes through Wi-Fi HaLow's network capabilities, ensuring that the system progressively refines itself over time and adapts to the changing dynamics of the urban environment."

4. Methodology

4.1. Network configuration

As previously discussed, both network technologies operate using IP protocols and the TCP/IP link layer, making the system agnostic to the specific technology used, whether for sending publication requests to the backend implemented in Thingspeak or for transmitting spectrogram images corresponding to the most recently detected second.

Each monitoring node employs an HT-HD01 dongle configured as a station, connected to a Raspberry Pi via Ethernet, acting as a bridge between the Wi-Fi HaLow network and the conventional IP network. Although Ethernet is the default connection method, the dongle supports three interfaces: USB Type-C, Ethernet, and 2.4 GHz Wi-Fi. The whole system can be powered autonomously by 18,650 Battery Shield V8 Mobile Power Bank 3 V/5 V batteries if you do not have access to your own power.

On the server side, a Heltec HT-H7608 access point is used. This device is equipped with an MT7628 MCU and advanced RF capabilities. It supports both Wi-Fi HaLow (IEEE 802.11ah) and 2.4 GHz Wi-Fi, allowing it to connect either to a conventional access point or a standard home router. Fig. 6 shows the final node used for the experimental tests where the connections and the elements that compose it can be seen, as well as the protective housing for outdoor installation.

For this system, the UDP protocol was selected for data transmission due to its simplicity and low overhead. This choice ensures that in the event of packet loss, the system remains operational without interruption, as UDP does not require acknowledgment of receipt. On the receiving side, spectrograms are stored in the database by species and in the order of arrival, allowing for real-time data accumulation and organization.Neither of the two network technologies employed (Wi-Fi HaLow or Ethernet) poses limitations in this context, as the transmitted data packets, each corresponding to a 98×50 Mel-spectrogram reshaped into a 1 × 4900 vector, are approximately 20 KB in size. The Simulink model utilizes a configured 'UDP Send' block with a buffer size of 32,768 bytes, which is more than sufficient to handle the transmission without fragmentation or delays.

For the network configuration, the Heltec HT-H7608 Wi-Fi HaLow router is connected to the server's main router, with a NAT

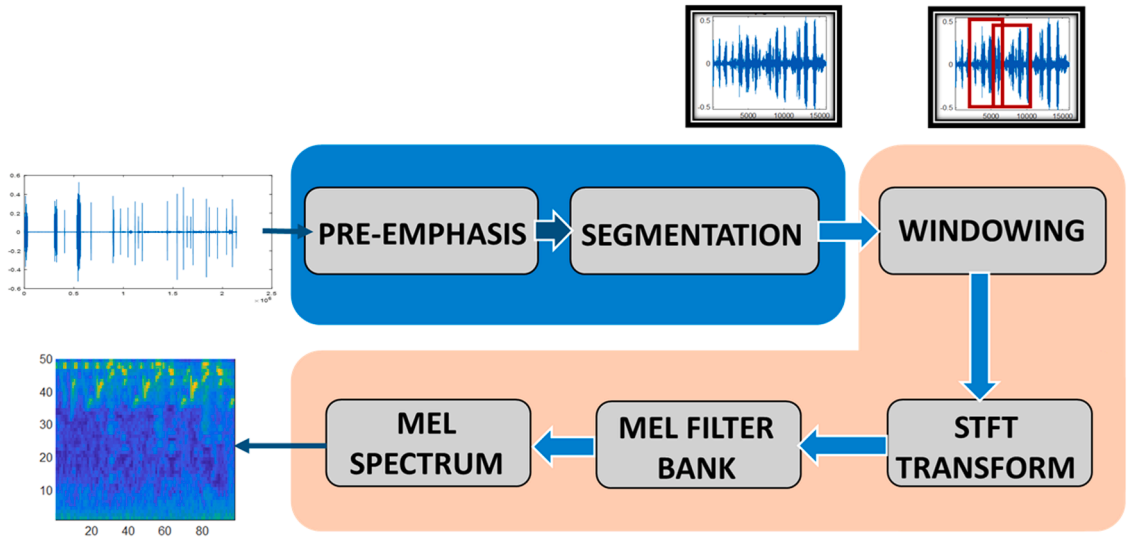


Fig. 5. Blue: preliminary phase of the training process. Orange: spectrogram extraction for each segment.

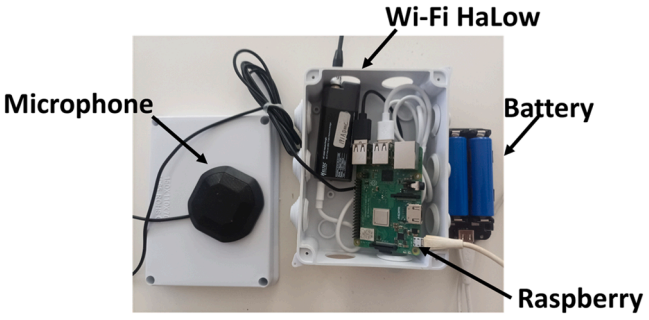


Fig. 6. Edge node implementation.

(Network Address Translation) setup implemented to interconnect the HaLow network with the conventional IP network. This router also functions as a DHCP server for devices connected to the HaLow network. On the node side, the Heltec HT-HD01 dongle is configured in bridge mode, enabling it to assign IP addresses to non-HaLow devices connected to it, such as the Raspberry Pi, which connects via its Ethernet interface in this case. As a result, the sensor node has both connections configured and can operate seamlessly using either of them.

The network protocols used are HTTP for the connection to Thingspeak and UDP for the envoi of the spectrograms. Both protocols are directly implemented in the two connection technologies. Thingspeak stores the received data in the cloud, which in this case consists of presence alerts for each species, while the spectrograms are stored on a local server using a directory structure specifically designed to support the subsequent training of other neural networks.

To receive the incoming UDP packets and convert them into 98 $\times$ 50 matrices and images, a data flow has been implemented in Node-RED. This flow listens on the specified UDP ports, performs reshaping of each packet into the required matrix format, and stores the resulting data into a structured format suitable for subsequent training tasks. Fig. 7 illustrates this data flow.

4.2. *Case study experiment*

To assess the system's performance under real-world conditions, a series of tests were conducted across different urban scenarios using the two proposed communication systems. The first scenario involves an urban park with conventional Wi-Fi coverage, while the second features a park equipped with a Wi-Fi HaLow access point located at a considerable distance, due to the absence of standard Wi-Fi infrastructure. These settings represent two common use cases frequently encountered in medium to large urban areas.

For detection, one-second audio clips sampled at 16,000 samples per second were used. To generate the spectrogram of each audio segment, 0.25 s frames were used for spectrum computation, with time hops of 0.01 s and a filter bank of 50 filters. The resulting spectrogram has a resolution of 98 $\times$ 50 pixels and serves as the input to the first convolutional layer, which highlights low-level features by generating a feature map. The audio files were sourced from the open-access database xeno-canto.org [43], which provides audio recordings in various formats and durations.

Moreover, in one of the test environments, a high density of rose-ringed parakeets (*Psittacula krameri*) was observed. This enabled an assessment of their potential ecological impact, particularly in terms of the presence or absence of other avian species, in comparison with locations where this invasive species is less abundant. For the duration of the test periods, the sensor nodes were mounted on poles or trees. The experiments were conducted over one-hour sessions on different days to evaluate the system's operational consistency and to validate the data through parallel human observation. Fig. 8 illustrates the selected deployment sites, along with the

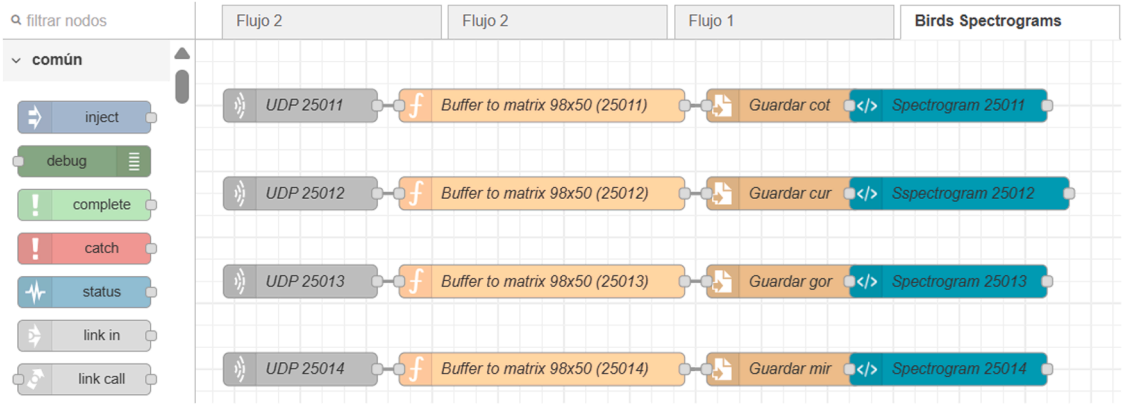


Fig. 7. Receiver data flow.

corresponding coverage conditions previously described. Zone b is managed by a HaLow Wi-Fi router located in a shopping centre about 200 m away.

The tests were conducted in these areas to evaluate both the system's communication range and the detection of the presence of different species, following the configurations shown in the figure. These areas represent two common and representative urban configurations.

4.3. Metrics

This section presents the types of measurements that will be obtained to evaluate the overall system. On one hand, the measurements related to the accuracy of sensor node detection are considered; on the other hand, the communication range measurements of existing technologies are compared against the propagation models presented in this work.

4.4. Wi-Fi HaLow models

Regarding the propagation of the Communications System in the scenarios seen above, a comparison is made between the actual range measurements obtained using Wi-Fi HaLow technology and the estimated coverage distances derived from simulations based on two signal propagation models: the Hata model and the Free Space Path Loss (FSPL) model.

The Hata model provides an empirical estimation of signal attenuation in urban environments, considering factors such as operating frequency, distance, and the height of both the transmitting and receiving antennas [44]. In contrast, the FSPL model represents an idealized scenario with no obstacles or reflections, offering a theoretical reference for maximum coverage under line-of-sight conditions.

Both models were used to calculate path loss as a function of distance under the same operational parameters (868 MHz frequency, antenna heights, and environmental conditions). The results were then compared with the actual communication distances observed during field tests using Wi-Fi HaLow devices. This comparison enables an evaluation of how well the theoretical models align with real-world behaviour and helps estimate the necessary safety margin for practical presence detection applications. Although the COST-231 Hata Model, for example, is more suitable for urban environments, due to the frequencies employed by Wi-Fi HaLow, it is not valid.

4.4.1. Hata model

This model is primarily applicable in urban environments, with a parameter that classifies based on urban density. Hata's model is an empirical formulation derived from Okumura's model and is used to estimate propagation loss in wireless communications, especially in urban, suburban, and rural environments.

It is recommended for frequencies between 150 MHz and 1500 MHz, making it suitable for HaLow Wi-Fi, which typically operates in the 800–928 MHz range.

Originally designed for mobile communications, the model requires a correction factor due to antenna height, as in this case the antennas are not mounted on communication towers. The correction factor is calculated as follows:

$$a_{hm} = \left(1,1 * log_{10}(f) - 0,7\right) * h_m - \left(1,56 * log_{10}(f) - 0,8\right) \tag{1}$$

where:

f = frequency in MHz (e.g., 868),

h_m = height of the receiving antenna (in meters), and

a_{hm} is expressed in decibels (dB).

This correction factor is subtracted in the total path loss calculation. The total path loss is calculated as follows:

$$Lu = 69,55 + 26,16 * log_{10}(f) - 13,82 * log_{10}(h_b) - a_{hm} + \left(41,9 - 6,55 * log_{10}(hb)\right) * log_{10}(d) \tag{2}$$

where:

f is the frequency in MHz (Wi-Fi HaLow),

Fig. 8. Selected areas: a) Wi-Fi coverture, b) Wi-Fi HaLow link.

d is the distance in kilometres (ranging from 50 m to 1000 m),
h_b is the height of the transmitting antenna in meters (e.g., a gateway mounted on a pole), and
h_m is the height of the receiving antenna in meters (e.g., a sensor).

In the case of suburban environments, this expression can be applied due to the lower presence of obstacles. It could be valid in our case in medium-sized park scenarios. In this case, the losses are given by:

$$L_{SU} = L_U - 2 * \left[log_{10}\left(\frac{f}{28}\right)\right]^2 - 5,4 \tag{3}$$

4.4.2. Free space path loss (FSPL) model

This model represents an ideal propagation scenario in which there are no obstacles, reflections, or multipath effects. It assumes a clear line-of-sight between the transmitter and the receiver. Although it does not account for real-world environmental conditions, the FSPL model serves as a theoretical reference for the maximum possible communication range at a given frequency.

The path loss in free space is calculated using the following expression:

$$PL = 20 * log_{10}(d) + 20 * log_{10}(f) + 32,44 \tag{4}$$

where:

d is the distance between transmitter and receiver in meters,
f is the frequency in MHz (e.g., 868 or 900 MHz for Wi-Fi HaLow), and
PL is the total path loss in decibels (dB).

This equation is derived from the *Friis* transmission formula and is commonly used in wireless communication systems to evaluate the baseline signal attenuation in free-space conditions. Because it ignores any environmental losses, its results are optimistic and should be interpreted as best-case coverage estimations.

4.5. Neural network

Before presenting the real-time results obtained in the field scenarios described earlier, we first introduce the outcomes of the training phase of the neural network using the curated bird song audio database. To evaluate the accuracy of the system, it is essential to consider four fundamental classifications: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Based on these, the following performance metrics are derived:

$$PAccurracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F1 - score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{8}$$

Accuracy represents the overall proportion of correct predictions made by the model across all classes, encompassing both avian vocalizations and background sounds. Precision, on the other hand, quantifies the proportion of true positive (TP) predictions among all instances classified as positive, thus reflecting the model's reliability in correctly identifying specific bird species when a detection occurs. The proportion of true negatives (TN) among all negative cases assesses the system's ability to accurately distinguish background environmental sounds from bird vocalizations.

Recall, also known as sensitivity or true positive rate, is a metric in machine learning that quantifies the proportion of actual positive instances correctly identified by a classification model, relative to the total number of true positives. Finally, the F1-score, also referred to as the F-score or F-measure, is a commonly used evaluation metric in data science and artificial intelligence. It represents the harmonic mean of precision and recall, and is particularly useful for assessing classification performance, especially in scenarios such as diagnostic testing where both false positives and false negatives carry significant weight. Collectively, these parameters indicate the model's capacity to accurately identify bird vocalizations while effectively differentiating them from background environmental noise.

5. Results

The CNN model deployed on the edge nodes was evaluated using a labeled test set consisting of 1259 samples. The dataset included four bird species: rose-ringed parakeet (*Psittacula krameri*), Eurasian blackcap (*Sylvia atricapilla heineken*), Spanish sparrow (*Passer hispaniolensis*), and common blackbird (*Turdus merula*), as well as a 'background' class representing non-bird ambient audio. For detection, one-second audio clips sampled at 16.000 Hz were used. To generate the spectrograms, each audio clip was segmented into

0.25 s frames with a time hop of 0.01 s, and a filter bank of 50 filters was applied. The resulting spectrograms had a resolution of 98 × 50 pixels and were used as input to the first convolutional layer, as illustrated in Fig. 9. These are the spectrograms that are transmitted by UDP to the server. The corresponding confusion matrix is presented in Fig. 10.

Following equations (5) through (8), the class-wise precision metrics presented in Fig. 11 were obtained. The overall classification accuracy reached 98.73 %, demonstrating the system's effectiveness for embedded bioacoustics detection.

In another subsequent test, the Canary Islands Chiffchaff (Phylloscopus canariensis) was included. In this case, the number of audio fragments is 1942. This model was used for scenario b). As before, the confusion matrix is shown in Fig. 12, and the corresponding performance metrics are presented in Fig. 13. In this case, the overall system accuracy slightly decreases to 96.69 %.

A detailed inspection of the confusion matrices (Figs. 10–12) reveals that the majority of errors arise from two specific situations. The *Chiffchaff* class shows slightly lower precision due to its very short and intermittent song, which occupies only a small fraction of the 1 s audio window used for analysis. Consequently, many windows contain mostly background noise, leading to false positive detections. In contrast, *Blackbird* and *Blackcap* exhibit strong acoustic similarity in both frequency content and temporal modulation, which explains their occasional mutual misclassifications. Apart from these cases, the remaining species maintain precision and recall above 97 %, confirming the robustness of the proposed model.

Figs. 14 and 15 show the detections in areas a) and b) as indicated in Fig. 8. In the first area, detections of the Eurasian Blackcap, the Common Blackbird, and the Spanish Sparrow can be observed. In contrast, in the second area, despite its proximity, a greater number of detections of the Rose-ringed Parakeet were recorded, along with detections exclusively of the Canary Islands Chiffchaff. At first

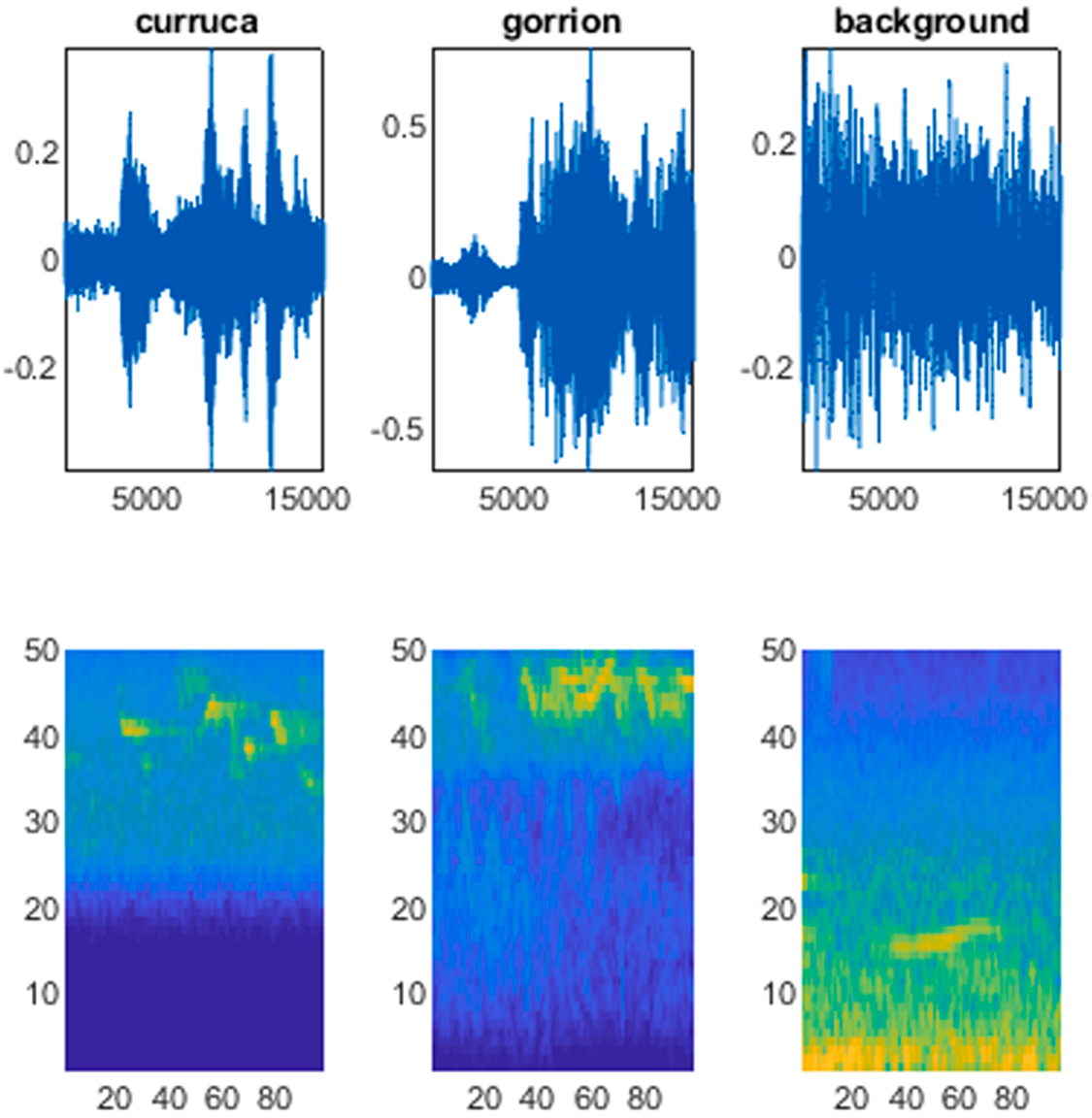


Fig. 9. Samples of the spectrograms obtained at a resolution of 98 × 50 pixels.

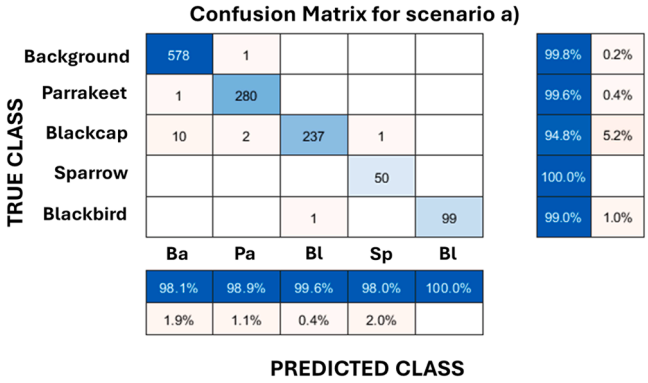


Fig. 10. Confusion matrix.

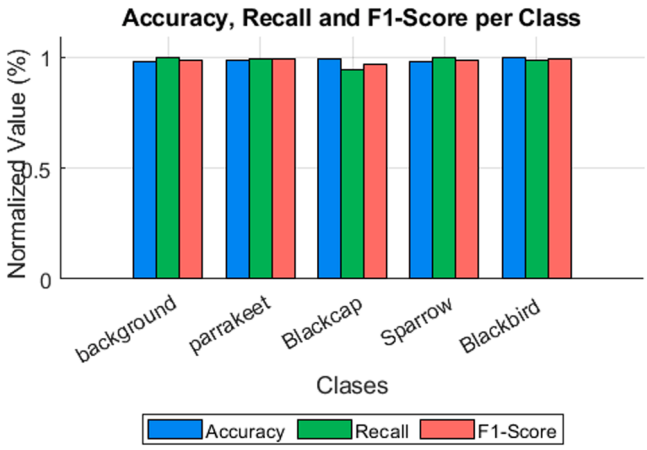


Fig. 11. Classification metrics per class.

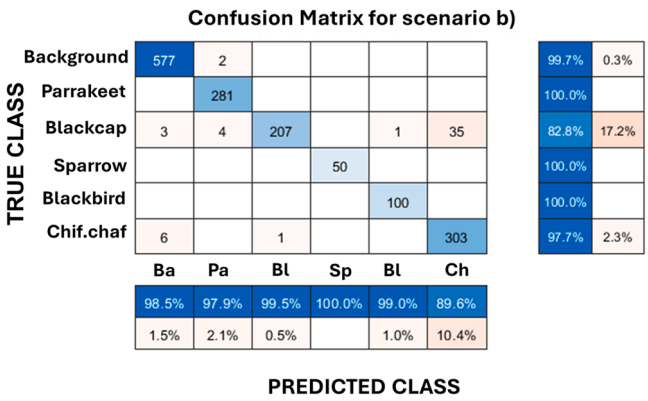


Fig. 12. Network confusion matrix trained for scenario b).

glance, this may be significant when analysing the potential influence of this species on other endemic species of the islands. These results are shown as they appear on ThingSpeak over periods of approximately one hour.

Fig. 16 shows the overall detections in both areas after two days of observations, to reflect the percentage of total detections. The influence of the Rose-ringed Parakeet on the presence of other species is clearly observable.

Power measurements on the Raspberry Pi 3+ with the Wi-Fi HaLow dongle indicate an average consumption of 5.3 W during active inference and ≈5.15 W in idle mode, corresponding to ∼1.06 J per processed spectrogram. Using a 10,000 mAh 18,650 battery, the node can operate for about 7 h under continuous load, confirming the platform's suitability for low-power, autonomous field

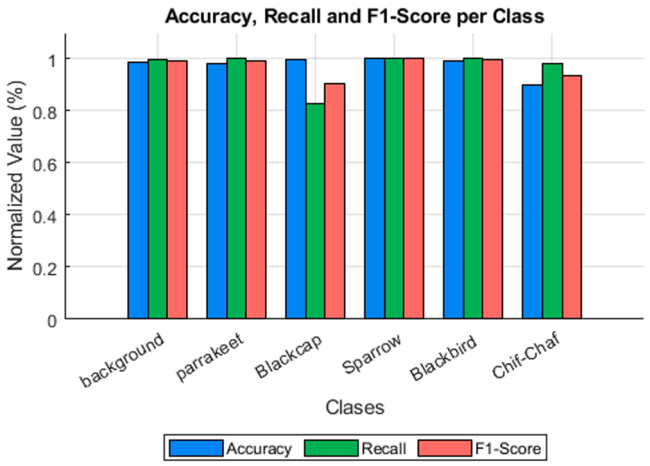


Fig. 13. Classification metrics per class for the network in scenario b).

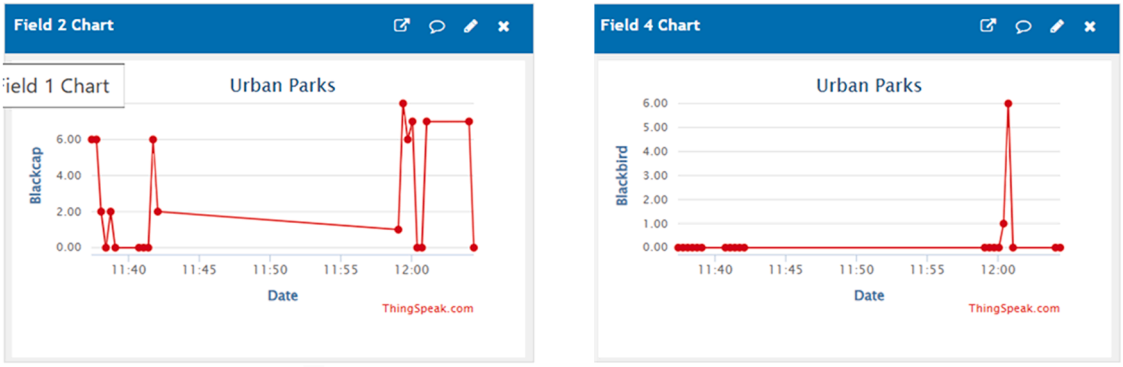


Fig. 14. Example of detections in area a).

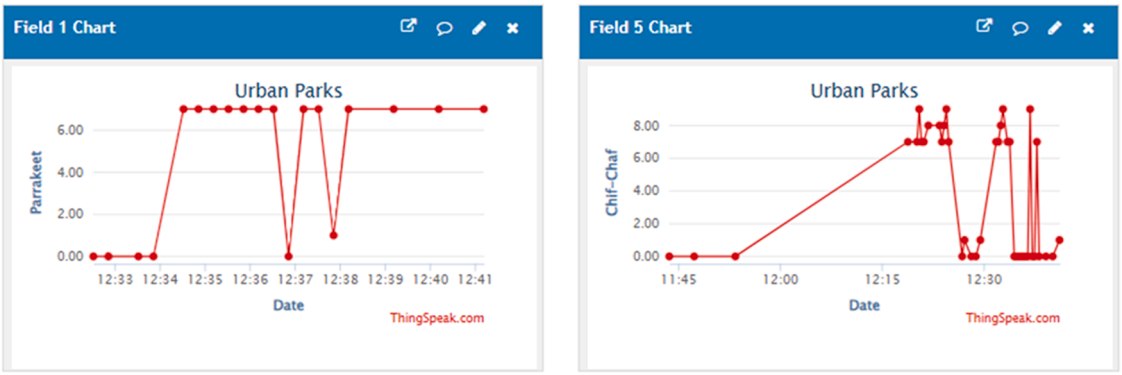


Fig. 15. Example of detections in area b).

deployments.

Regarding the implemented Wi-Fi HaLow network, a comparison has been made with the propagation models discussed previously. In the presented scenario, the coverage was as expected; however, depending on the distance, the bit rate decreased from 3 Mbps to 1 Mbps. This data rate proved to be more than sufficient for transmitting the spectrograms of the detections, as well as for sending alerts to Thingspeak. Fig. 17 shows the comparative between Hata model, Hata sub-urban model and Free space. The figure includes the measurements made with the System in the form of points. As can be observed, although the scenario is located in an area with few obstacles, the measurements are closer to the urban model than to the suburban one. The main consequence of this is increased system

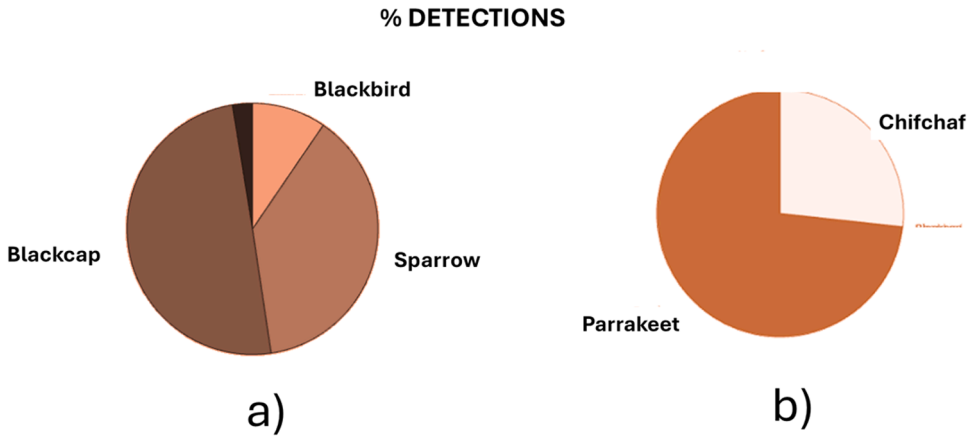


Fig. 16. Percentage of detections after two days in areas a) and b).

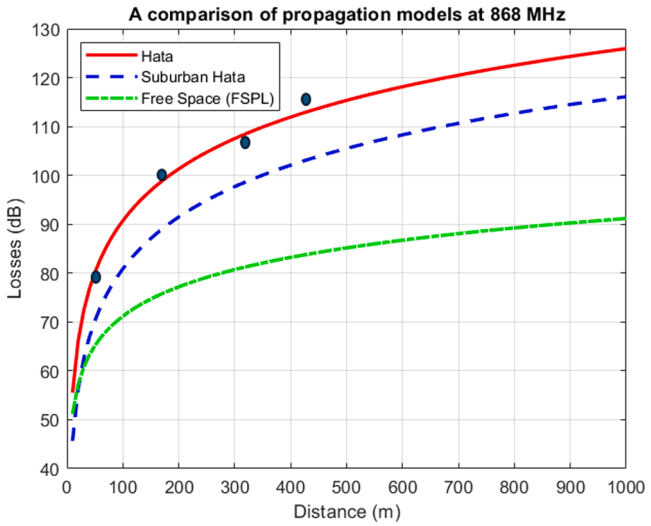


Fig. 17. Comparison between three models and real measurements taken at four points.

latency and the appearance of packet loss. This justifies the use of UDP, which does not require a persistent connection. As for the losses, they do not pose any issues in the short or medium term. The latency levels obtained are quite variable due to the adaptation of the system to the different types of modulation, but the average in the scenario seen in Fig. 8 is 250 ms, with maximums of 3000 ms.

Although latency measurements varied between 250 ms and 3 s, this range remains acceptable for real-time invasive species monitoring, where detection and alerting operate on multi-second timescales. The latency variability observed does not affect the timeliness or reliability of the alert system.

The UDP-based transmission scheme was experimentally validated with two nodes sending data at low rates ($\approx$ 1 packet every few seconds). Under these conditions, the measured packet loss did not exceed 12 %, which had no impact on the effective detection rate since the number of valid detections remains proportional to the total transmissions. Additionally, a theoretical worst-case analysis of channel occupancy as a function of node count was performed to estimate scalability limits. Even in conservative scenarios, the estimated channel load remains moderate, confirming that UDP transmission is adequate for the proposed low-traffic acoustic monitoring system.

Regarding the network limits, they were calculated through the percentage of channel load for cases with 1 Mbps and 6 Mbps physical layer rates, with a number of nodes ranging from 1 to 50, and assuming a distribution of UDP spectrogram packet transmissions with inter-arrival times varying between 10 and 20 s (a rather extreme case, since detections usually do not occur this frequently). The calculation was carried out assuming randomized transmissions from each node.

First, the average arrival rate λ is calculated as:

$$\lambda = \frac{1}{15} = 0.06667 s^{-1} \tag{9}$$

The duration of a packet transmission is:

$$D = \frac{S}{R} \tag{10}$$

where S is the length of the packet and R is the bitrate of the network. So the instantaneous probability that a node is transmitting is given by:

$$\rho = \lambda \bullet D \tag{11}$$

And the percentage of network occupancy is:

$$U = N \bullet \rho \tag{12}$$

where N is the number of nodes.

With these calculations, the comparative graph in Fig. 18 is obtained

As observed, when transmitting packets of 120,000 KB from 10 nodes, the network approaches saturation. This behavior is analyzed under the assumption of highly frequent transmission intervals, representing a worst-case scenario rather than typical operational conditions. Table 3 presents the results of our system compared to previous studies in terms of architecture, connectivity, and on-site versus cloud processing, highlighting that our approach supports real-time operation and on-device spectrogram generation.

6. Discussion

This work demonstrates the feasibility of monitoring bird activity using a low-cost, low-power system embedded in a Raspberry Pi, capable of operating autonomously by adapting its communication techniques to environmental conditions. In urban environments with Wi-Fi coverage, the system utilizes the Raspberry Pi's built-in network interface, while in areas requiring extended range, a Wi-Fi HaLow module is integrated.

Unlike other similar approaches, this system does not rely on audio recording or streaming. Instead, detections are performed in real time, as both detection and classification processes are executed at the EDGE layer of the system architecture. This eliminates the need for high-bitrate data transmission. The five-layer CNN was selected to Goulãonce accuracy and efficiency under embedded constraints. On the Raspberry Pi 4, the model requires <2 MB of memory and achieves inference times below 30 ms per 1 s spectrogram, ensuring real-time operation. Compared with heavier networks such as MobileNetV2 or EfficientNet-B0, our compact design enables efficient deployment while leaving processing capacity available for additional modules. The versatility of the Raspberry Pi platform also allows seamless integration of new communication interfaces, such as low-power **Wi-Fi HaLow**, for distributed or remote acoustic monitoring applications.

High overall accuracy and macro-averaged F1 scores demonstrate the model's effectiveness in classifying bird vocalizations while minimizing misclassifications. Beyond its application for wildlife monitoring, the system also offers valuable insights into the broader dynamics of urban ecosystem biodiversity. Moreover, it exhibits a performance level comparable to expert observations in detecting invasive species, such as the rose-ringed parakeet (*Psittacula krameri*), in real time making it a highly useful tool to support scientific research and conservation efforts.

The observations presented regarding the invasive parakeet and native species are preliminary and illustrative. The primary contribution of this work is the implementation of a robust bioacoustic monitoring platform capable of collecting comprehensive datasets over time. While statistical analysis of species interactions (e.g., chi-square tests, *t*-tests) is not performed here, the system enables such analyses in future studies, supporting long-term ecological research and biodiversity assessments.

Notably, both network technologies utilized in this study support TCP/IP traffic, thereby enabling the transmission of data packets containing one-second spectrograms for each detection event. This constitutes a significant advantage over other low-power wide-area network (LPWAN) technologies such as LoRa or NB-IoT, which either do not support such transmissions or require complex preprocessing steps to enable them. The adoption of Wi-Fi HaLow, owing to its intrinsic characteristics, permits dynamic modulation scheme adjustments based on environmental conditions such as distance. It adapts responsively to parameters like received signal strength indicator (RSSI) or signal-to-noise ratio (SNR). Consequently, the bit rate decreases, while latency and packet loss tend to increase as distance grows. Nevertheless, within this type of system, such variations are generally not deemed critical.

To prevent database saturation, the transmission of spectrograms is regulated by a timer mechanism, which mitigates the excessive load that would otherwise result from transmitting data for every individual detection over prolonged periods.

Moreover, the storage of spectrograms for future model training offers context-specific acoustic data, allowing the system to adapt over time to evolving urban soundscapes (e.g., construction activity, vehicular traffic, or human presence). This enhances the system's capacity to generalize across the diverse acoustic conditions characteristic of real-world urban environments. This strategy also obviates the need for a segmentation phase, which would otherwise be necessary when working with existing audio datasets containing segments of variable length and acoustic properties, thus yielding substantial savings in both time and computational resources, as illustrated in Fig. 19.

The continuous learning process was tested using newly acquired audio segments to evaluate the efficiency of the adaptive data handling and preprocessing stages. As shown in Fig. 19, these experiments resulted in a measurable reduction in CPU load and processing time, confirming the feasibility of real-time operation on the Raspberry Pi platform. While the retraining stage was not yet

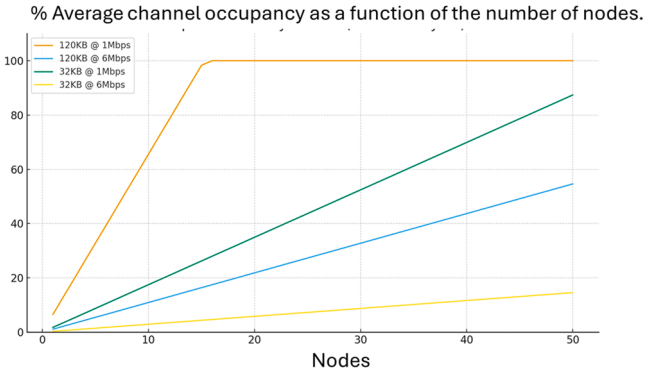


Fig. 18. Percentage of channel occupancy based on the number of nodes.

Table 2
Comparison between 802.11 and 802.11ah.

Feature	Wi-Fi HaLow (802.11ah)	Conventional Wi-Fi (802.11n/g)
Operating frequency	Sub-1 GHz (US 902–928 MHz, EU 863–868 MHz)	2.4 GHz (802.11b/g/n), 5 GHz (802.11n/ac)
Maximum range	Up to 1 km	50–100 m (typical)
Maximum data rate	~0.15 Mbps to 35 Mbps	Up to 600 Mbps (802.11n)
Power consumption	Very low (optimized for IoT devices)	Medium to high
Obstacle penetration	High (effective through walls)	Moderate
Latency	Moderate	Low
Common applications	IoT, smart agriculture, smart cities	Web, video streaming, general networking
Performance in urban environments	Superior coverage and range	Susceptible to interference

performed, these tests validate the core functionality of the continuous learning loop and its readiness for future iterative updates.

As can be observed, the overall accuracy of the system does not increase significantly; however, when the model is adapted to each specific environment, its performance improves within those contexts. This is due to the system's ability to account for the unique acoustic characteristics of each environment, including background noise and local sound patterns, which enhances detection and classification accuracy at the local level.

In comparison with other systems presented in the literature, this is the only approach that enables the automatic generation of a database containing audio segments of appropriate duration directly from the EDGE layer. Furthermore, by performing processing at the edge, the system achieves significantly lower latency, allowing for real-time operation. Table 2 presents a comparative analysis with other systems discussed before.

7. Conclusions

To summarize, the proposed system represents a significant contribution to urban biodiversity monitoring. By leveraging low-cost, low-power devices capable of edge processing, it enables real-time detection and classification while reducing latency and dependence on cloud infrastructure. Despite relying on edge computing, the system achieves classification accuracy that is fully comparable to state-of-the-art approaches.

Moreover, the use of TCP/IP-based network technologies enables the transmission of small data packets, including images and spectrograms of detected species. This contributes to building a growing database of pre-segmented and pre-classified spectrograms, facilitating future research and retraining processes.

This design also enhances scalability and adaptability. The system can be retrained iteratively to suit new environments, improving detection accuracy over time. As a result, it significantly reduces processing time and eliminates the need for manual segmentation of audio recordings, a common limitation in current bioacoustics datasets, where song detection and segmentation are often done manually before training.

A novel addition is the integration of Wi-Fi HaLow, a technology particularly suited for IoT and Smart City deployments due to its low power consumption, extended range, and compatibility with widely used devices such as the Raspberry Pi. Unlike alternative IoT technologies like LoRa or NB-IoT, Wi-Fi HaLow maintains full support for the TCP/IP protocol stack, ensuring reliable data transmission and real-time alerts, despite a slight reduction in bit rate.

8. Future works

Currently, efforts are underway to adapt the system to other technologies such as LTE or NB-IoT to achieve similar functionalities. This involves an interoperability phase between the protocols used by these systems and the possibility of transmitting spectrogram

Table 3
Comparison between different similar studies in terms of architecture, connectivity and where they perform the processing.

Study	Architecture	Need for coverage	Communications technology	Real time/ spectrogram
[14]	Pipeline. Soundscape recordings previously collected	No	None	No/No
[7]	Based in mobile recordings	Yes	4G	No/No
[10]	Pre-processing on sensor nodes and post-processing in the cloud	Yes	Wi-Fi	Yes/No
[17]	On-site processing with camera traps and sending alarms via Wi-Fi	Yes	Wi-Fi	Yes/No
This Work	On-site processing using microphone and Rasperry Pi and sending alarms via Wi-Fi or Wi-Fi HaLowLoRa	Not in the case of Wi-Fi HaLow up to 1 Km	Wi-Fi Wi-Fi HalLow	Yes/Yes

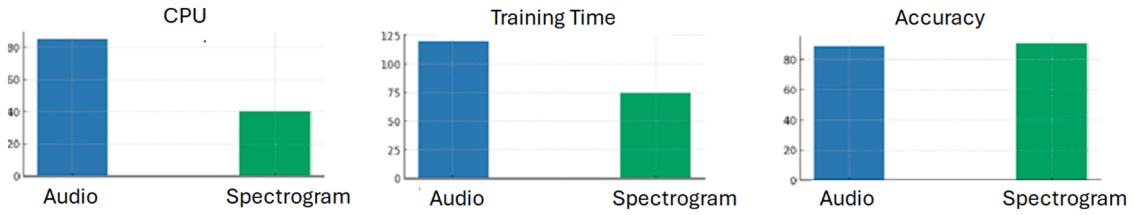


Fig. 19. Relative CPU usage, training time and accuracy using spectrograms from the detections.

fragments over TCP/IP. Such an approach would enable greater coverage without the need to deploy additional access points, particularly in urban areas where 4G or 5G coverage is readily available.

Additionally, with the development of new devices such as Arduino boards or FPGAs, it would be possible to create sensor nodes with this type of connectivity, embedding the entire system similarly to how it is integrated into the Raspberry Pi. It is also possible to improve the implementation of the current system by using other HaLow Wi-Fi devices such as NRC7292 that have even lower consumption and using antennas with higher gain.

It should be noted that the propagation model comparison presented here is theoretical and does not account for real-world interference. However, in Europe, the 868 MHz frequency band is primarily used for LoRa communications and is not commonly congested in urban environments. Empirical evaluation under interference conditions will be addressed in future work.

Finally, the system is scalable to a wide range of environment-related detections, including biodiversity monitoring and the integration of a greater number of sensors. This can be done in a straightforward manner, making the system both scalable and flexible without requiring major modifications.

CRediT authorship contribution statement

Francisco A. Delgado-Rajó: Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Investigation, Conceptualization. **Carlos M. Travieso-González:** Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Ruyman Hernández-López:** Writing – review & editing, Software, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Reports a relationship with that includes:. Has patent pending to. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M. Goulão, L. Bandeira, B. Martins, A.L. Oliveira, Training environmental sound classification models for real-world deployment in edge devices, Discov. Appl. Sci. 6 (4) (2024) 166, https://doi.org/10.1007/s42452-024-05803-7.
[2] Boussard, H., et al. (2021). Towards citizen science for smart cities: A framework for a collaborative game of bird call recognition based on internet of sound practices. *arXiv preprint.* arXiv:2103.16988.
[3] J.F. Chace, J.J. Walsh, Urban effects on native avifauna: a review, Landsc. Urban Plan. 74 (1) (2006) 46–69, https://doi.org/10.1016/j.landurbplan.2004.08.007.
[4] M.C. Zwart, et al., Monitoring of mammal and bird species in an urban ecological park using eDNA metabarcoding, Urban Ecosyst. (2024), https://doi.org/10.1007/s11252-024-01557-7.
[5] J.K. Schneider, et al., NDVI and vegetation volume as predictors of urban bird diversity, Sci. Rep. (2025) 15, https://doi.org/10.1038/s41598-025-96098-0.
[6] R. Kumar, et al., Artificial intelligence for sustainable urban biodiversity: a framework for monitoring and conservation, arXiv preprint (2024) arXiv:2501.14766.
[7] Rovithis, E. & Moustakas, N. & Vogklis, K. & Drossos, K. & Floros, A.. (2021). Towards citizen science for Smart cities: a framework for a collaborative game of bird call recognition based on internet of sound practices. 10.48550/arXiv.2103.16988.
[8] D. Stowell, et al., Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge, Methods Ecol. Evolut. 10 (3) (2019) 368–380, https://doi.org/10.1111/2041-210X.13103.
[9] A. Thakur, et al., Real-time bioacoustic classification with edge computing for urban wildlife monitoring, Sensors 23 (2) (2023) 456, https://doi.org/10.3390/s23020456.
[10] T. Cinkler, N. Kristóf, C. Simon, R. Vida, H. Rajab, Two-phase sensor decision: machine-learning for bird sound recognition and vineyard protection, IEEE Sensors J. (2021), https://doi.org/10.1109/JSEN.2021.3134817.
[11] F. Delgado-Rajo, A. Melian-Segura, V. Guerra, R. Perez-Jimenez, D. Sanchez-Rodriguez, Hybrid RF/VLC network architecture for the Internet of things, Sensors 20 (2) (2020) 478, https://doi.org/10.3390/s20020478.
[12] J.S. Gomes, et al., LoRa-based communication for wildlife acoustic monitoring: a performance evaluation, Ad Hoc Netw. 135 (2022) 102985, https://doi.org/10.1016/j.adhoc.2022.102985.

[13] T. Andreassen, A. Surlykke, J. Hallam, Semi-automatic long-term acoustic surveying: a case study with bats, Ecol. Inform. 21 (2014) 13–24, https://doi.org/10.1016/j.ecoinf.2013.12.010.
[14] J. LeBien, M. Zhong, M. Campos-Cerqueira, J.P. Velev, R. Dodhia, J. Lavista Ferres, T.M. Aide, A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network, Ecol. Inform. 59 (2020) 101113, https://doi.org/10.1016/j.ecoinf.2020.101113.
[15] A. Selin, J. Turunen, J.T. Tanttu, Wavelets in recognition of bird sounds, EURASIP J. Adv. Signal Process 2007 (2007) 51806, https://doi.org/10.1155/2007/51806.
[16] S. Hu, Y. Chu, Z. Wen, G. Zhou, Y. Sun, A. Chen, Deep learning bird song recognition based on MFF-ScSEnet, Ecol. Indic. 154 (2023) 110844, https://doi.org/10.1016/j.ecolind.2023.110844. ISSN 1470-160X.
[17] D. Velasco-Montero, J. Fernández-Berni, R. Carmona-Galán, A. Sanglas, F. Palomares, Reliable and efficient integration of AI into camera traps for smart wildlife monitoring based on continual learning, Ecol. Inform. 83 (2024) 102815, https://doi.org/10.1016/j.ecoinf.2024.102815. ISSN 1574-95412815.
[18] D. Strubbe, E. Matthysen, Establishment success of invasive ring-necked and monk parakeets in Europe, J. Biogeogr. 36 (12) (2009) 2264–2278, https://doi.org/10.1111/j.1365-2699.2009.02174.x.
[19] R.R Noumida, Deep learning-based automatic bird species identification from isolated recordings, in: 2021 8th International Conference on Smart Computing and Communications (ICSCC), 2021, pp. 252–256, https://doi.org/10.1109/ICSCC51209.2021.9528234.
[20] S. Kahl, CM. Wood, M. Eibl, H. Klinck, BirdNET: a deep learning solution for avian diversity monitoring, Ecol. Inform. 61 (2021) 101236, https://doi.org/10.1016/j.ecoinf.2021. ISSN 1574-9541.
[21] D. Stowell, M. Wood, Y. Stylianou, H. Glotin, Bird detection in audio: a survey and a challenge, in: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing, MLSP, 2016, pp. 1–6, https://doi.org/10.1109/MLSP.2016.7738875.
[22] T.M. Aide, C. Corrada-Bravo, M. Campos-Cerqueira, C. Milan, G. Vega, R. Alvarez, Real-time bioacoustics monitoring and automated species identification, PeerJ 1 (2013) e103, https://doi.org/10.7717/peerj.103.
[23] P. Jančovič, M. Kküer, Automatic detection and recognition of tonal bird sounds in noisy environments, EURASIP J. Adv. Sign. Proc. 2011 (2011) 982936, https://doi.org/10.1155/2011/982936.
[24] N. Priyadarshani, S. Marsland, I. Castro, Automated birdsong recognition in complex acoustic environments: a review, J. Avian Biol. 49 (2018) jav–01447, https://doi.org/10.1111/jav.01447.
[25] M. Goitia-Urdiain, T. Sauras-Yera, GA. Llorente, Eudald Pujol-Buxó, software-dependent biases in the recognition of di- and tri-syllabic bird songs can create false interpretations of bird abundance and singing activity, Ecol. Inform. 79 (2024) 102397, https://doi.org/10.1016/j.ecoinf.2023.102397. ISSN 1574-9541.
[26] B. Chandu, A. Munikoti, K.S. Murthy, G. Murthy, C. Nagaraj, Automated bird species identification using audio signal processing and neural networks, in: 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, 2020, pp. 1–5, https://doi.org/10.1109/AISP48273.2020.9073584.
[27] M. Arowolo, W. Aaron, A. Kugbiyi, U. Eteng, D. Iloh, C. Aguma, A. Olagunju, Integrating AI enhanced remote sensing technologies with IOT networks for precision environmental monitoring and predicative ecosystem management, World J. Adv. Res. Rev. 23 (2024) 2156–2166, https://doi.org/10.30574/wjarr.2024.23.2.2573.
[28] J. Segura-Garcia, S. Sturley, M. Arevalillo-Herraez, J.M. Alcaraz-Calero, S. Felici-Castell, E.A. Navarro-Camba, 5G AI-IoT system for bird species monitoring and song classification, Sensors 24 (11) (2024) 3687, https://doi.org/10.3390/s24113687W.
[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018.
[30] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: Proceedings of the International Conference on Machine Learning (ICML), 2019.
[31] J.C Shi, Q. Zhang, Y. Li, L. Xu, Edge Computing: vision and challenges, IEEE Internet Things J. 3 (5) (2016) 637–646, https://doi.org/10.1109/JIOT.2016.2579198. Oct.
[32] A. Khan, M. Faheem, R.N. Bashir, C. Wechtaisong, M.Z. Abbas, Internet of Things (IoT) assisted context aware fertilizer recommendation, IEEE Access 10 (2022) 129505–129519, https://doi.org/10.1109/ACCESS.2022.3228160.
[33] A. Prinz, V. Taank, V. Voegeli, E. Walters, A novel nest-monitoring camera system using a Raspberry Pi micro-computer, J. Field Ornithol. 87 (2016), https://doi.org/10.1111/jofo.12182. http://doi.
[34] Youngblood,M. A Raspberry pi-based, RFID-equipped birdfeeder for the remote monitoring of wild bird populations. 2019/01/02. doi: https://doi.org/10.1080/03078698.2019.1759908.
[35] Z. Huang, et al., TinyChirp: bird song recognition using TinyML models on low-power wireless acoustic sensors, in: 2024 IEEE 5th International Symposium on the Internet of Sounds (IS2), Erlangen, Germany, 2024, pp. 1–10, https://doi.org/10.1109/IS262782.2024.10704131.
[36] V. Bonilla, B. Campoverde, S.G. Yoo, A systematic literature review of LoRaWAN: sensors and applications, Sensors 23 (20) (2023) 8440, https://doi.org/10.3390/s23208440.
[37] A. Diane, O. Diallo, E.H.M. Ndoye, A systematic and comprehensive review on low power wide area network: characteristics, architecture, applications and research challenges, Discov. Internet Things 5 (2025) 7, https://doi.org/10.1007/s43926-025-00097-6.
[38] B.S. Chaudhari, M. Zennaro, S. Borkar, LPWAN technologies: emerging application characteristics, requirements, and design considerations, Future Intern. 12 (3) (2020) 46, https://doi.org/10.3390/fi12030046.
[39] L. Vangelista, Frequency shift chirp modulation: the LoRa modulation, IEEE Sign. Proc. Lett. 24 (12) (2017) 1818–1821, https://doi.org/10.1016/j.ufug.2017.08.014.
[40] Y. Kim, J. Jang, H. Lee, Performance evaluation of wi-fi HaLow in IoT environments, Sensors 23 (6) (2023) 2954, https://doi.org/10.3390/s23062954.
[41] Wireless Broadband Alliance, *Wi-Fi HaLow for IoT*. WBA White Paper, Retrieved from, https://wballiance.com/resource/wifi-HaLow-iot, 2024.
[42] The MathWorks Inc, MATLAB (Version R2024a), The MathWorks, Inc, Natick, Massachusetts, 2024.
[43] xeno-canto, Sharing bird sounds from around the world. Retrieved May 13, 2020, from, https://www.xeno-canto.org/, 2020.
[44] Y. Singh, Comparison of Okumura, Hata and COST-231 models on the basis of path loss and signal strength, Int. J. Comput. Appl. 59 (2012) 37–41, https://doi.org/10.5120/9594-4216.

Further reading

[45] E. Kristiani, C.-T. Yang, C.-Y. Huang, P.-C. Ko, H. Fathoni, On construction of sensors, edge, and cloud (iSEC) framework for Smart system integration and applications, IEEE Internet Things J. 8 (1) (2021) 309–319 1 Jan.1, doi: https://doi.org/10.1109/JIOT.2020.3004244.
[46] Z. Huang, Y. Chen, L. Wang, Smart reference evapotranspiration using Internet of Things and hybrid ensemble machine learning approach for precision irrigation, Comput. Electro. Agric. 204 (2023) 107554, https://doi.org/10.1016/j.compag.2023.107554.
[47] A.A. Khan, et al., Context aware evapotranspiration (ETs) for saline soils reclamation, IEEE Access 10 (2022) 110050–110063, https://doi.org/10.1109/ACCESS.2022.3206009.
[48] Kumar, G.P. Hancke, Energy efficient environment monitoring system based on the IEEE 802.15.4 standard for low cost requirements, IEEE Sens. J. 14 (8) (2014) 2557–2566 Aug, doi: https://doi.org/10.1109/JSEN.2014.2313348.
[49] E. Rovithis, N. Moustakas, K. Vogklis, K. Drossos, A. Floros, Towards citizen science for smart cities: a framework for a collaborative game of bird call recognition based on internet of sound practices, ArXiv (2021), https://doi.org/10.48550/arXiv.2103.16988.abs/2103.16988.

[50] S. Dian Handy Permana, K. Bayu Yogha Bintoro, Implementation of constant-Q transform (CQT) and Mel spectrogram to converting Bird's sound, 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Purwokerto, Indonesia, 2021, pp. 52–56, doi: https://doi.org/10.1109/COMNETSAT53002.2021.9530779.
[51] A.&. Seferagic, S.&. Kerkhove, Le& Tian, J.&. Famaey, A.&. Munteanu, I.&. Moerman, J. Hoebeke, E. De Poorter, Performance evaluation of IEEE 802.11ah networks with high-throughput bidirectional traffic, Sensors 18 (2018) 325, https://doi.org/10.3390/s18020325.
[52] P. Tryjanowski, et al., Bird diversity in urban green space: a large-scale analysis of differences between parks and cemeteries in Central Europe, Urban Forest. Urban Green. 27 (2017) 264–271 rights and content, doi: https://doi.org/10.1016/j.ufug.2017.08.014Get.