IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 May 2025, accepted 12 June 2025, date of publication 18 June 2025, date of current version 25 June 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3580620*

RESEARCH ARTICLE

# Balancing Complexity and Performance of Machine Learning Models for Avian Pests Sound Detection in Agricultural Environments

**MICHELINE KAZENEZA[1,2], (Member, IEEE), ANNA SERGEEVNA BOSMAN[2], (Member, IEEE), DESTINY KWABLA AMENYEDZI[1,3], (Student Member, IEEE), DAMIEN HANYURWIMFURA[1], EMMANUEL NDASHIMYE[4], (Member, IEEE), AND ANTHONY VODACEK[3], (Senior Member, IEEE)**

[1]African Centre of Excellence in Internet of Things, University of Rwanda, Kigali, Rwanda
[2]Department of Computer Science, University of Pretoria, Pretoria 0002, South Africa
[3]Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623, USA
[4]School of ECE, Carnegie Mellon University Africa, Kigali, Rwanda

Corresponding author: Micheline Kazeneza (kazeneza_221030473@stud.ur.ac.rw)

This work was supported in part by the icipe-World Bank Financing under Agreement D347-3A, in part by the World Bank-Korea Trust Fund for Partnership for Skills in Applied Sciences, Engineering and Technology (PASET)-Regional Scholarship and Innovation Fund (RSIF) under Agreement TF0A8639, and in part by the University of Pretoria through PASET-RSIF Research Funds.

**ABSTRACT** Agricultural pest control traditionally relies on inefficient visual inspections. Acoustic monitoring offers a promising alternative by analyzing pest-specific sounds. While effective, implementing acoustic monitoring in agricultural settings faces practical constraints, particularly the limited computational resources available in remote farming environments. This necessitates optimized machine learning (ML) solutions for low-power edge devices. This study evaluates ML models for bird pest detection on resource-constrained platforms. We evaluated convolutional neural networks (CNNs), recurrent neural networks (RNNs), and traditional ML models by comparing standalone and knowledge-distilled versions of EfficientNetB0 and gated recurrent unity (GRU) against EfficientNetB4, Long short-term memory (LSTM), MobileNetV2, LightGBM, and support vector machine (SVM). Analysis revealed significant performance variations across computational requirements. LightGBM achieved 98% accuracy with minimal resources (8,500 parameters, 7KB, 0.6ms inference), demonstrating good efficiency. SVM (97% accuracy) and distilled GRU (86% accuracy) also showed favorable performance-to-resource ratios. Knowledge distillation substantially enhanced the accuracy of EfficientNetB0 (from 73% to 98%) and modestly improved GRU (from 84% to 86%). We examined platform compatibility across computing tiers, discovering that while high-performance edge devices (Jetson Nano, Raspberry Pi 4) support all studied models effectively, microcontrollers require specialized approaches. Advanced microcontrollers (such as ESP32-S3 and STM32H7) can accommodate optimized implementations, while highly constrained platforms (such as Arduino Nano) require TinyML techniques. This research contributes 1) an on-farm audio dataset, 2) comprehensive cross-model evaluation metrics, and 3) deployment optimization strategies for acoustic pest detection systems in resource-constrained agricultural environments.

**INDEX TERMS** Acoustic monitoring, bird pest detection, deep learning, knowledge distillation, machine learning.

## I. INTRODUCTION

Traditional farming relies heavily on visual inspection and manual interventions, which are time-consuming and expensive. During field observations in Rwanda, numerous

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

bird species were identified that caused substantial damage to grain, vegetable, and fruit crops [1], [2] alongside beneficial birds. Modern automated acoustic detection systems offer improved agricultural management through targeted pest control that minimizes disturbance to beneficial birds, potentially reducing crop damage while enhancing productivity and sustainability [1]. For implementation in resource-constrained agricultural environments, these solutions must be deployable on inexpensive edge devices that meet the affordability requirements of rural farmers.

Given the uncertainty about which approach would provide the optimal performance-to-cost ratio, we conducted a comprehensive evaluation of diverse machine learning models and feature extraction techniques. Our assessment compares recurrent neural networks (GRU, LSTM) analyzing raw audio waveforms [3], CNN architectures (EfficientNet, MobileNet) processing spectrograms [4], and classical ML approaches (LightGBM and SVM) trained on mel frequency cepstral coefficient (MFCCs) [5], [6]. This systematic comparison aims to identify the most efficient and effective solutions for bird pest detection that can be practically deployed in low-resource agricultural settings. To further optimize these solutions for edge deployment, we apply knowledge distillation techniques [7], [8] that transfer expertise from larger models (EfficientNetB4, LSTM) to their smaller counterparts (EfficientNetB0, GRU), enhancing performance within the computational constraints of edge computing [9]. Through this systematic evaluation of models, feature extraction methods, and knowledge distillation, this study establishes a foundation for developing cost-effective bird pest detection systems deployable on resource-constrained edge devices.

The rest of the paper is structured as follows: Section II summarises related work. Section III details the materials and methods used to conduct the research. Section IV presents and discusses the experimental findings. Finally, Section V concludes the study and proposes some directions for future work.

## II. RELATED WORK

Acoustic monitoring has become increasingly important for agricultural pest management, with various approaches demonstrating effectiveness across different contexts. Mahjoub et al. [10] established acoustic interventions as viable tools for bird control, showing that targeted sound approaches can significantly influence avian behavior. Brandes [11] provided foundational techniques for automated sound recording and classification of bird vocalizations that inform modern detection systems. These studies validated acoustic approaches for pest management, though their focus remained primarily on detection capabilities rather than deployment efficiency evaluation in resource-constrained environments.

Specialized systems like BirdNet [12] represent significant advances in avian bioacoustics. Based on CNN architecture adaptations, BirdNet can distinguish among 984 bird species with high accuracy by analyzing spectrogram data from bird vocalizations [12], [13]. However, BirdNet's dataset shows significant geographical bias, with data primarily from North America and Europe while largely overlooking more than 2,700 bird species across Africa. Furthermore, while demonstrating the potential of deep learning for bird sound analysis, BirdNet's complexity (approximately 30 million parameters) creates substantial barriers for implementation in agricultural settings where connectivity, energy, and computational resources are limited.

The integration of IoT technologies has transformed agricultural monitoring by enabling remote oversight of crop conditions. However, IoT deployments face significant resource constraints, particularly for computationally intensive applications. Researchers have addressed these challenges through specialized architectures like MobileNetV2 [14], which incorporate depthwise separable convolutions to reduce computational requirements while maintaining acceptable performance. Similarly, EfficientNet variants [15] employ compound scaling methods to optimize the balance between network depth, width, and resolution, making them suitable for resource-constrained deployments.

For audio processing in limited-resource environments, several machine learning approaches have proven effective. RNN architectures, particularly LSTM and GRU models [16], [17], excel at capturing temporal dependencies in audio data. GRUs offer parameter efficiency advantages over LSTMs, making them suitable for deployment in constrained settings [18]. CNN-based models like EfficientNet variants have demonstrated strong performance in sound classification and audio event detection [19], while MobileNet architectures provide efficient alternatives for real-time audio processing applications [20].

Traditional machine learning approaches offer further efficiency benefits. LightGBM [21] provides an optimized gradient-boosting framework capable of handling tabular data with minimal computational overhead. Support vector machines (SVMs) [22] perform effectively even with limited training samples and have shown success in various audio classification tasks [6], [23].

Knowledge distillation (KD) techniques [7], [8] provide a promising approach for optimizing deep learning models in resource-constrained environments. By training smaller student models to emulate the behavior of larger, more complex teacher models, KD enables the transfer of sophisticated capabilities while significantly reducing computational requirements. This approach is particularly valuable for agricultural deployment, where balancing detection accuracy with resource limitations is essential for practical implementation.

Our research builds upon these foundations to address the specific challenges of bird pest detection in resource-constrained agricultural settings. Through the systematic evaluation of various models and optimization techniques, we provide practical insights for implementing effective acoustic monitoring systems in environments with limited computational resources, connectivity, and energy

**TABLE 1. Comparison between existing systems and proposed approach.**

| Existing Systems | Proposed Approach |
|---|---|
| BirdNet [12] claims to be 'comprehensive' but has geographical bias, with data primarily from North America and Europe, overlooking 2700+ African bird species. Furthermore, it requires high computational resources (27M parameters). | Specifically targets pest bird detection with significantly reduced computational demands (as low as 8.5K parameters) while better representing relevant local species. |
| Acoustic monitoring systems [10], [11], [24] emphasize detection capabilities without addressing deployment efficiency. | Systematically evaluates resource efficiency metrics critical for practical agricultural implementations. |
| Deep learning architectures [14], [15] provide general optimization approaches for mobile/IoT deployment. | Applies and compares multiple model types (CNN, RNN, traditional ML) specifically for bird pest detection. |
| Knowledge distillation [7], [25] is established as a theoretical approach for model compression. | Implements and evaluates KD techniques specifically for acoustic pest detection, demonstrating practical efficiency gains. |
| Audio processing models [16], [19] focus on general sound classification without agricultural specialization. | Tailors audio processing pipelines for bird pest sounds in complex agricultural acoustic environments. |
| Traditional ML approaches [21], [22] demonstrated for general classification tasks. | Adapts traditional ML methods specifically for bird pest detection with comprehensive comparisons to deep learning approaches. |

availability. Table 1 presents a comparison between existing systems and the proposed approach, highlighting the key differences and innovations of our work.

## III. MATERIALS AND METHODS

This section outlines the research approach followed and the techniques employed. Subsection III-A details the data collection process. Subsection III-B explains the data preprocessing, Subsection III-C describes the classification procedure, and the experimental setup of this study is discussed in Subsection III-D.

### A. DATA COLLECTION AND DESCRIPTION

The dataset comprises audio recordings from a wheat field at the University of Rwanda's CAVM campus in Busogo. The study site adjoined roads, housing, and a restaurant, creating a complex acoustic environment of human activity, traffic sounds, bird pests, and nonpest vocalizations. AudioMoth microphone devices [26] captured 45-60 second recordings at 48kHz during daylight hours, strategically positioned throughout the field. Figure 1(a) shows a deployed AudioMoth microphone, while Figure 1(b) displays the geographical location ($-1.556, 29.549$) of the study area. The field recordings exhibit high acoustic complexity, characterized by overlapping sound sources within similar frequency ranges. One example is shown in Figure 1(c). This recording demonstrates multiple bird vocalizations overlapping with environmental sounds across the 1-10 kHz range. Rwanda's dense population and rich avian diversity create complex acoustic environments with simultaneous multi-source audio signals occupying overlapping frequency domains.

### B. DATA PREPROCESSING

Appropriate data preparation is essential for accurate model performance. Our study employed different preprocessing techniques tailored to each model type. RNN-based models (LSTM/GRU) used raw audio data, CNN models (EfficientNet/MobileNet) utilized spectrograms, while LightGBM and SVMs processed extracted MFCCs. Each preprocessing strategy forms an integral part of the model's deployment pipeline, with its computational demands factored into the total efficiency assessment.

Subsection III-B1 explains the data labeling process, Subsection III-B2 presents the audio segmentation techniques, and Subsection III-B3 details the spectrogram calculation and generation.

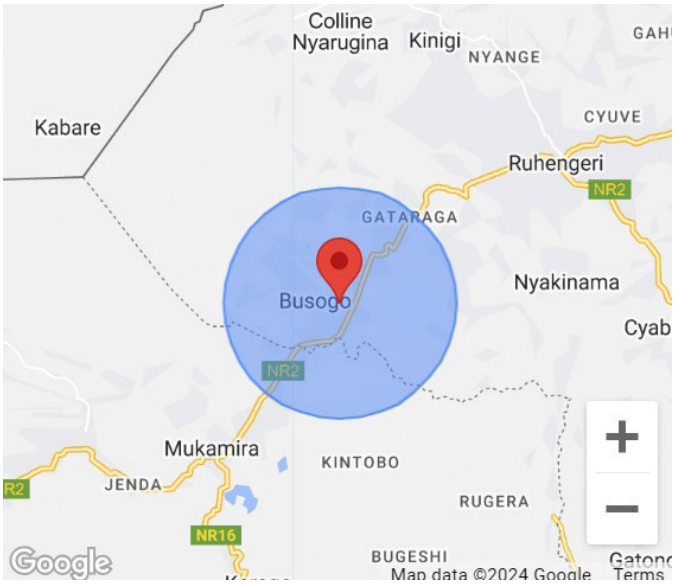
#### 1) LABELING

To analyze the acoustic environment of the farm under study, the research utilized Kaleidoscope Pro software, which specializes in audio clustering and labeling [27]. The initial phase involved clustering, an iterative process that categorized the collected farm sounds into distinct groups based on their acoustic properties. Kaleidoscope Pro is designed to identify the presence of sound in a recording [28], subsequently searching through all provided recording files for similar sounds and outputting the resulting clusters. Each cluster was initially assigned a default label by the software, such as `cluster.00`, `cluster.01`, `cluster.02`,..., `cluster.11`, `cluster.12`, etc. During this iterative process, we identified misclassifications where sounds were erroneously grouped with dissimilar audio types. To address these anomalies, the team manually reassigned such sounds to more appropriate clusters that shared similar sonic characteristics. These preliminary labels were then subjected to a refinement process. Through the manual review, the research team analyzed the content of each sound cluster, subsequently assigning more descriptive and contextually appropriate labels. This was achieved by adding two labels, "bird-pest" and "no-pest" to the descriptive `.CSV` file generated by Kaleidoscope Pro in the `MANUAL ID` column to accurately reflect whether the sounds were from bird pests or other environmental sources. Notably, avian sounds associated with benign bird species were placed in the "no-pest" category.

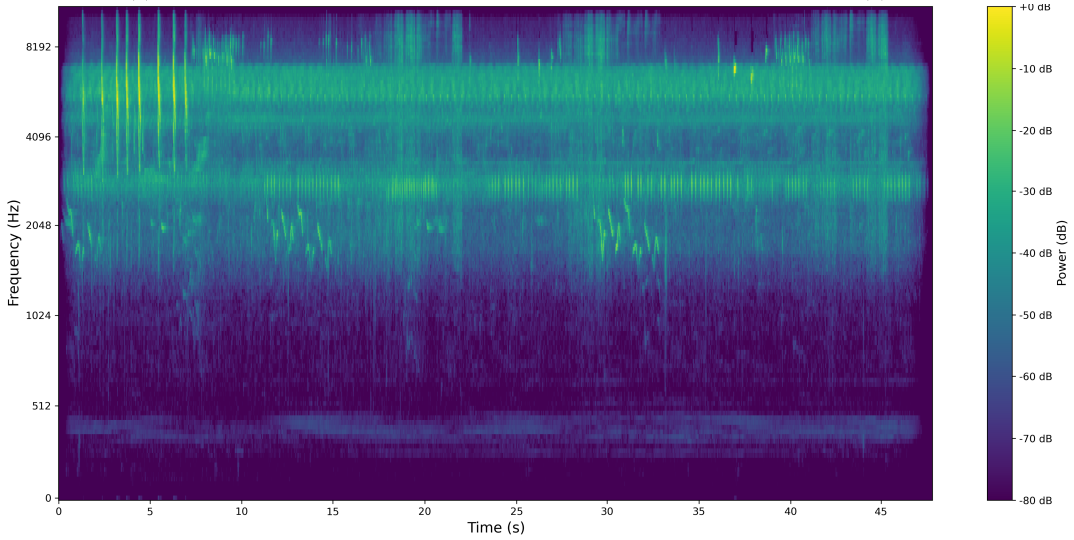
#### 2) DATASET PREPARATION

The original field recordings, spanning 45-60 seconds each, contained diverse, overlapping sounds from multiple sources. To enhance classification precision, these recordings were

(a)




(b)




(c)

**FIGURE 1.** **Data collection site. (a) AudioMoth was placed in a wheat farm at Busogo campus, Rwanda, and (b) the location of Busogo on Google Maps, where the audio data was collected. Lat −1.556, Lon 29.549. (c) the spectrogram of one recorded sample, highlighting the dataset's complexity characterized by multiple overlapping sounds from different sources.**

segmented into 1-5 second clips that isolated specific acoustic events with their environmental context. This duration aligns with typical bird vocalization lengths and was automatically determined by Kaleidoscope Pro software during sound clustering. Initial segmentation yielded 9,070 audio samples: 8,163 classified as ''bird-pest'', and 907 as ''no-pest''. The dataset was deliberately structured as a binary classification problem, ''bird-pest'' versus ''no-pest'', to align with the current step of the system's application requirements. This binary classification serves as an initial development phase. It addresses the immediate practical objective of determining when intervention is necessary. This approach provides actionable information by distinguishing pest birds from other environmental sounds. Future iterations of the approach may be expanded to a more refined classification framework that differentiates between pest birds, beneficial birds, and environmental noise.

The ''bird-pest'' category encompasses recordings of crop-damaging avian species, including common waxbills (*Estrilda astrild*), common nightingales (*Luscinia megarhynchos*), red fire finches (*Lagonosticta senegala*), streaky-headed seedeaters (*Sporophila lineola*), yellow bishops (*Euplectes capensis*), yellow-fronted canaries (*Serinus canaria*), Chubb's cisticola (*Cisticolidae*), village weavers (*Ploceus cucullatus*), and bronze mannikins (*Spermestes cucullata*). Conversely, the ''no-pest'' category comprises

beneficial birds like pied-winged swallows (*Hirundo rustica*) that prey on crop-threatening insects, human activities (conversations, play, music), vehicular noise, and non-pest animals such as dogs. To address the dataset imbalance (8,163 audio samples of ''bird-pest''versus 907 as ''no-pest'') and enhance model robustness, we supplemented the field recordings with environmentally compatible sounds from public repositories. Specifically, we incorporated sounds from the UrbanSound8k dataset,[1] that exhibit acoustic characteristics similar to our target environment, expanding the non-pest class to 8,161 samples. The resulting composite dataset comprises 16,324 audio samples with a balanced distribution: 8,163 bird pest vocalizations and 8,161 non-pest environmental sounds, providing a robust foundation for model training and evaluation.

#### 3) SPECTROGRAM CALCULATION AND GENERATION

The spectrogram calculation and generation within audio preprocessing offers insights into audio signal's temporal and spectral dynamics [29], [30]. This involves the use of the short-time Fourier transform, a mathematical transformation adept at segmenting the audio signal into discrete time intervals while concurrently unveiling its frequency components [30]. The resultant spectrogram manifests as a visual representation, depicting the evolution of frequency energy across time. Spectrogram calculation and subsequent visual representation of audio emerge as a useful tool for extracting nuanced information from raw audio signals, and enabling the application of powerful models designed for image processing, such as CNN-based models and others [31]. Figure 2 shows fifteen spectrogram images of the audio samples used in the study. It is visually evident from the spectrograms that each sample is unique, showcasing the distinctive audio signatures of the samples.

### C. MACHINE LEARNING FOR PEST IDENTIFICATION

In this section, the ML algorithms applied in this study are detailed. Subsection III-C1 summarizes the proposed methodology. Subsection III-C2 briefly discusses ML models used. Subsection III-C3 describes the KD process.

#### 1) METHODOLOGY OVERVIEW

The goal of this study was to systematically explore and compare a wide array of ML models and training methodologies, to identify the most efficient and effective ML models for real-time pest detection on farms, ensuring that the chosen models are both high-performing and viable for deployment in settings with constrained computational resources. Figure 3 summarises the investigative approach utilized in this study.

We explored two distinct modeling strategies to determine the most appropriate models: first, using KD with complex deep learning models to enhance the performance of smaller deep learning models, and second, directly training selected models without any intermediate processes.

Referring to Figure 3, the training process involved different steps. Initially, we explored KD techniques. We trained two large deep learning models, EfficientNetB4 and LSTM, to be ''teacher'' models for two compact deep learning models, EfficientNetB0 and GRU, respectively. This process involved two phases: (i) pre-training the teacher models on the target task, and (ii) training smaller student models, EfficientNetB0 and GRU, to emulate the teachers' performance. The student models were trained in two contexts: (i) directly under the guidance of their respective teachers, and (ii) independently, to evaluate their performance without distilled knowledge. This approach allowed us to assess the standalone capabilities of smaller models, while also showcasing the benefits of KD in enhancing model performance under computational constraints.

Additionally, to provide a comprehensive efficiency comparison, we trained simpler ML models, specifically MobileNetV2, SVM, and LightGBM. Various features were explored in model training to identify the most efficient methods. LightGBM and SVM models were trained using MFCC features as input, while spectrogram data was used for the CNN-based models (EfficientNets and MobileNets). Lastly, for the RNN-based models (GRU and LSTM), raw audio data was provided directly without extensive preprocessing. This approach is particularly advantageous for deployment on edge devices with limited resources because preprocessing operations like spectrogram or MFCC extraction require additional computational overhead and memory.

#### 2) MACHINE LEARNING MODELS OVERVIEW

Our model selection strategy was driven by the need to evaluate different trade-offs between accuracy, inference speed, and memory footprint for bird pest detection on edge devices. Each architecture offers distinct advantages for deployment in resource-constrained agricultural environments.

*Light Gradient Boosting Machines (LightGBM):* LightGBM is an efficient gradient-boosting decision tree-based framework, excelling in classification tasks across various domains, including bird sound recognition, EEG signal analysis, and medical diagnoses [6], [32], [33], [34], [35]. LightGBM success is attributed to its ability to handle structured numerical and categorical data [21], making it suitable for tasks such as audio signal processing using MFCCs [5]. We selected LightGBM for its fast inference times and minimal memory requirements compared to deep learning models, which are critical advantages for deployment on constrained IoT devices.

*Support Vector Machines (SVM):* SVM is a supervised max-margin model known for its effectiveness in audio classification, as demonstrated in various studies [6], [36], [37], [38]. In this study, SVM used MFCCs as inputs. SVMs provide an important baseline and often outperform neural

[1]UrbanSound8K - Urban Sound Datasets. Available at: https://urbansounddataset.weebly.com/urbansound8k.html

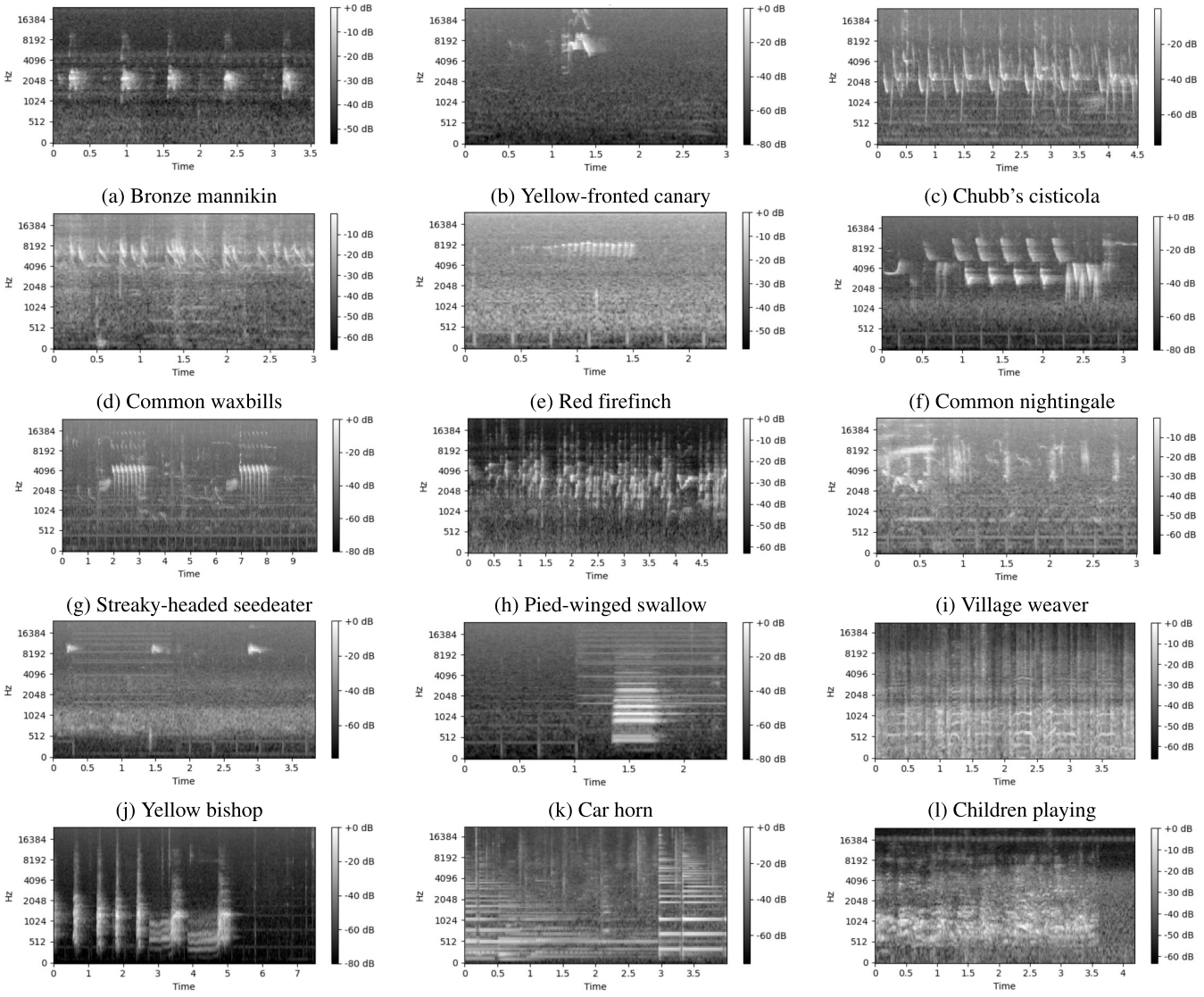



**FIGURE 2.** **Spectrograms providing visual representations of acoustic patterns from analyzed pest bird species plus ambient environmental sounds.**

networks in scenarios with limited training data [39]. Their simple computational structure and small memory footprint make them particularly attractive for microcontroller deployment where RAM is severely limited [40].

*EfficientNetB4:* EfficientNetB4 is a CNN model that employs a balanced sizing approach to improve network structure dimensions [15], enhancing accuracy and efficiency [41]. EfficientNetB4 is relatively complex, with 19 million total parameters. We chose it as a teacher model to train a simpler, lightweight EfficientNetB0 (4.2 million parameters) for deployment in resource-limited environments.

*EfficientNetB0:* EfficientNetB0, the smallest EfficientNet variant, offers an optimal balance between performance and resource efficiency, making it suitable for devices with limited computing power [15], [42]. While both transformer-based models [43] and specialized audio models like BirdNET [12] deliver superior classification results, their substantial computational requirements (with BirdNET requiring 27 million parameters and over 36MB storage) prevent them from providing affordable solutions for pests detection in farms, where farmers prioritize cost-effective and sustainable technologies. EfficientNetB0 maintains 90-99% of larger models' accuracy with significantly fewer resources, making the more compact architecture better suited for our application.

*MobileNetV2:* MobileNetV2 improves upon its predecessor [14] by connecting narrower network layers through a flipped design approach [44]. This architecture improves parameter efficiency while preserving detection accuracy, making it well-suited for real-time monitoring applications where power consumption is critical [45]. Both EfficientNet

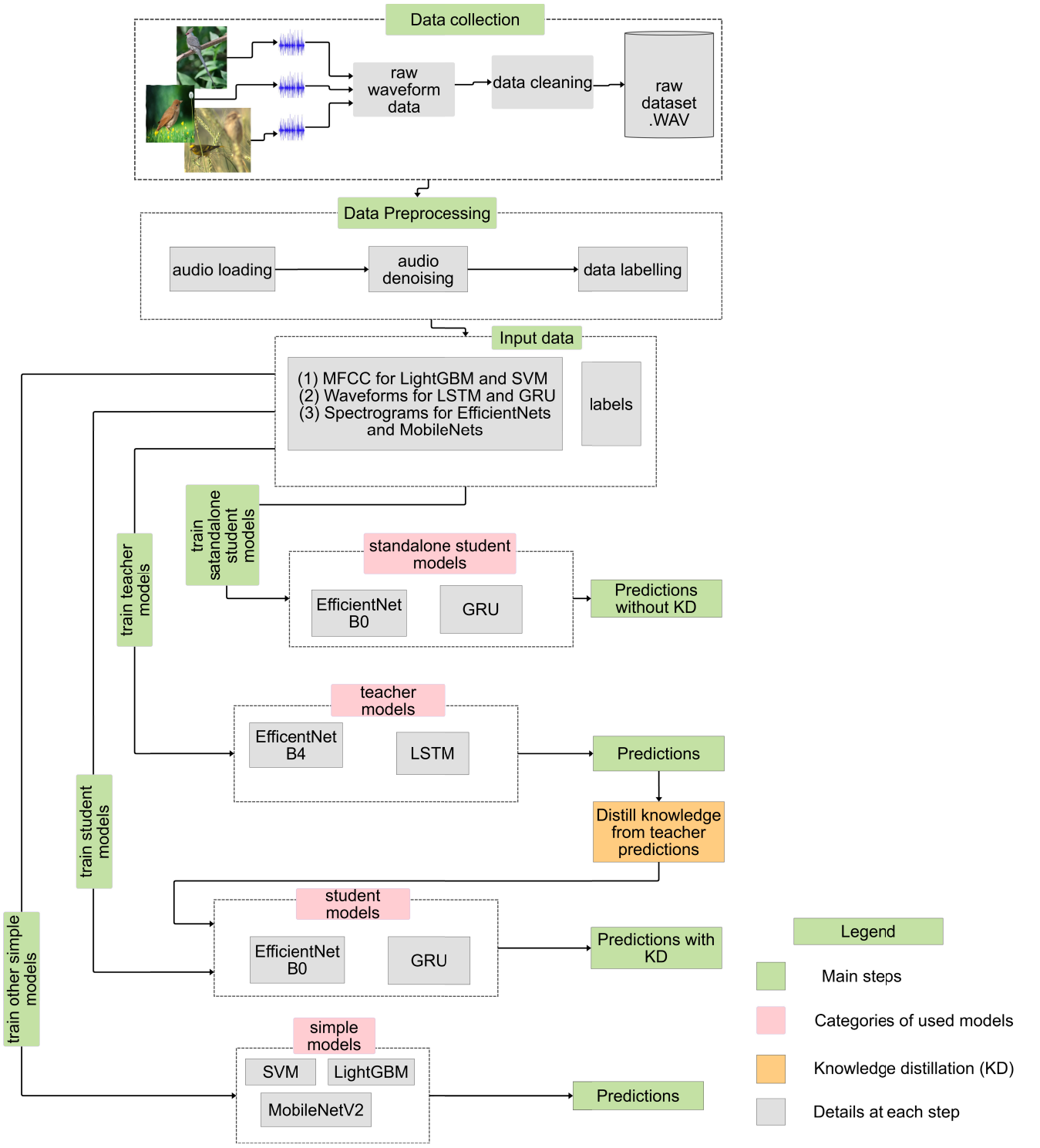



**FIGURE 3.** The overall structure of the proposed approach.

and MobileNet architectures were chosen over alternatives such as ResNet, Vision Transformers, or transformer-based audio models due to their optimized efficiency on resource-constrained devices. Unlike specialized audio models, such as PANNs [46] or WavLM [47] that prioritize accuracy at the expense of computational demands, our selected models balance performance with practical deployment requirements.

*Long Short-Term Memory (LSTM):* LSTM, a specialized recurrent neural network with gating mechanisms [48],

effectively learns temporal patterns in sequential data. While transformer-based systems such as Wav2Vec 2.0 [49] achieve superior audio recognition at the cost of substantial computational resources, LSTMs provide a more efficient solution for agricultural deployments while still capturing temporal patterns in bird vocalizations [50]. In our approach, we specifically selected LSTM to serve as the teacher model for training a more lightweight GRU architecture.

*Gated Recurrent Unit (GRU):* GRUs are gated RNNs that excel in resource-limited environments by processing raw audio with minimal preprocessing [18]. They use reset and update gates to control information flow [51], simplifying LSTM's gating structure. GRUs require approximately 25% fewer parameters than equivalent LSTMs while maintaining similar audio classification accuracy [50], making them suitable for RAM-constrained embedded systems. In our approach, we trained GRU both independently and as a student model to LSTM. Unlike approaches that prioritize accuracy regardless of computational cost (such as ensemble methods used by BirdCLEF competition winners [52]), our methodology emphasizes practical deployment viability while maintaining acceptable accuracy for bird pest detection in agricultural settings.

### 3) KNOWLEDGE DISTILLATION (KD)

KD serves as a method for reducing the computational complexity of machine learning models by transferring knowledge from a larger, more complex model to a smaller, more efficient one [53]. KD allows small models to benefit from the sophisticated capabilities of their larger counterparts without the associated computational overhead [53], [54]. During the KD process, the ''dark knowledge'' [55] is transferred from a more complex teacher network to a simpler student network. The ''dark knowledge'' refers to the implicit information contained in the probability distribution of a model's predictions [55]. This probability distribution serves as a source of ''soft targets'' for the student. When training the student model, the soft (probabilistic) targets are used to guide the learning process, helping the smaller model replicate the performance of the larger model more accurately than if the student was trained directly on the binary targets of the dataset. Figure 4 illustrates dark knowledge, where a teacher model predicts the class ''bird-pest'', with a probability of 0.68, and ''no-pest'' with a probability of 0.32.

Predicted probabilities (0.68 and 0.32) reveal not only the most likely class, but also the confidence level of the teacher model regarding its prediction. The probability distribution of the teacher's predictions is obtained using the modified softmax function, adjusted by incorporating the parameter $T$:

$$p_i = Z_{i,T} = \frac{\exp\left(\frac{Z_i}{T}\right)}{\sum_j \exp\left(\frac{Z_j}{T}\right)} \tag{1}$$

where $T$ is the constant and represents a *temperature* parameter, typically set to 1, and $Z_i$ corresponds to the $i$-th output of the model. Adjusting $T$ to higher values results in a smoother distribution of probabilities across classes. Figure 5 illustrates the effect of increasing $T$. A higher value of $T$ produces softer probability distributions, which present a more pronounced learning signal for the student model. A lower $T$ (approaching zero) results in probabilities close to 1 and 0, which may mask useful information during the learning process. Conversely, a higher $T$ allows the student model to learn from a richer, more nuanced set of probabilities, reflecting a broader spectrum of the underlying distribution.

The training of the student model integrates two types of losses or objective functions: the distillation, or soft loss ($L_{\text{soft}}$) and the student, or hard loss ($L_{\text{hard}}$) [56]. $L_{\text{soft}}$ is the binary cross-entropy loss between soft targets and soft predictions. Soft targets are the probability distribution output by the teacher model, while soft predictions are the output probability generated by the student model at a high temperature $T > 1$, shared by both the teacher and the student. $L_{\text{hard}}$ is the binary cross-entropy loss between hard targets and hard predictions. Hard targets are the actual labels of the dataset, and hard predictions are the categorical outputs of the student model at $T = 1$. The combined loss $L(x, W)$, incorporating both ($L_{\text{soft}}$) and ($L_{\text{hard}}$), is used to train the student model, effectively balancing the emulation of the teacher model's behavior with accurate independent predictions:

$$L(x, W) = \alpha L_{\text{hard}} + (1 - \alpha) L_{\text{soft}}, \tag{2}$$

where $L_{\text{hard}}$ is the hard loss, $L_{\text{soft}}$ is the soft loss, and $\alpha$ is a hyperparameter that balances the contribution of the hard and soft losses to the total loss. Figure 6 illustrates the construction of ($L(x, W)$).

The binary farm sound classification approach presented in this study forms the foundation for a more comprehensive multinomial system. In this binary framework, we replace the softmax function with a sigmoid function, reducing the output layer to a single neuron that maps real-valued numbers to the (0, 1) interval for probability interpretation, where 0 represents a negative class and 1 represents a positive class. To further implement the knowledge distillation in this binary classification context, a temperature parameter $T$ is incorporated into the sigmoid function for both positive and negative classes:

$$\hat{y}_{T,pos} = \delta\left(\frac{z}{T}\right) = \frac{1}{1 + \exp(-z/T)}, \tag{3}$$

$$\hat{y}_{T,neg} = \delta\left(\frac{\bar{z}}{T}\right) = 1 - \frac{1}{1 + \exp(-z/T)} = \frac{\exp(-z/T)}{1 + \exp(-z/T)}, \tag{4}$$

where $z$ is the logit output from the model, and $T = 7$. Figure 7 shows the range of values of the sigmoid function for positive and negative classes with different values of $T = \{1, 3, 5, 7\}$.

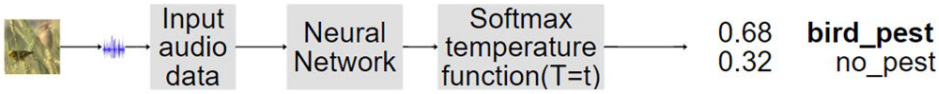



**FIGURE 4.** Dark knowledge output by the softmax temperature function while training the model at the temperature parameter $T = t$.

$$\text{T} = 1: \begin{bmatrix}0.99\\0.01\end{bmatrix} \begin{bmatrix}bird\_pest\\no\_pest\end{bmatrix} \quad \text{T} = 2: \begin{bmatrix}0.94\\0.06\end{bmatrix} \begin{bmatrix}bird\_pest\\no\_pest\end{bmatrix} \quad \text{T} = 3: \begin{bmatrix}0.89\\0.11\end{bmatrix} \begin{bmatrix}bird\_pest\\no\_pest\end{bmatrix}$$

$$\text{T} = 4: \begin{bmatrix}0.82\\0.18\end{bmatrix} \begin{bmatrix}bird\_pest\\no\_pest\end{bmatrix} \quad \text{T} = 5: \begin{bmatrix}0.73\\0.27\end{bmatrix} \begin{bmatrix}bird\_pest\\no\_pest\end{bmatrix} \quad \text{T} = 7: \begin{bmatrix}0.68\\0.32\end{bmatrix} \begin{bmatrix}bird\_pest\\no\_pest\end{bmatrix}$$

**FIGURE 5.** Probability distribution generated using the softmax temperature function for the positive class at a parameter $T = \{1, 2, 3, 4, 5, 7\}$.

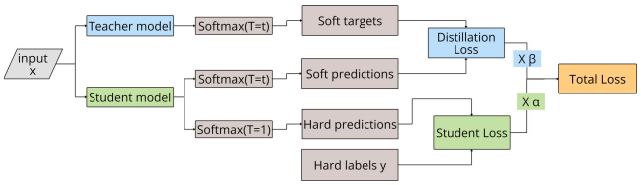



**FIGURE 6.** Flow chart showing how the total loss, employed to optimize the student architecture, is calculated during the KD process. Here, $\beta = 1 - \alpha$. Colors are used for readability.

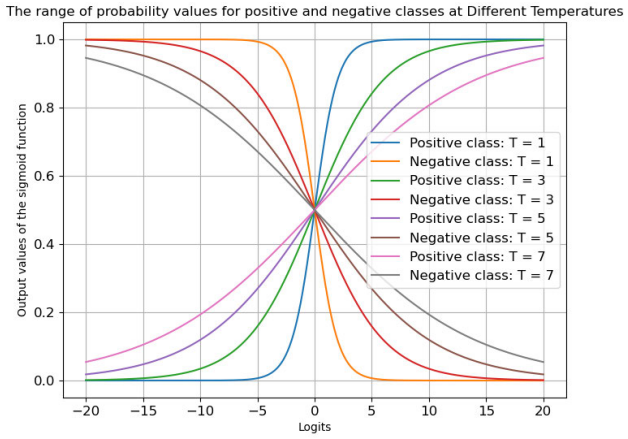



**FIGURE 7.** Sigmoid function for both positive and negative classes with T values of 1, 3, 5, and 7. The range of output values is shown for each T parameter.

### D. EXPERIMENTAL SETUP

Experiments were conducted using Jupyter Notebook version 7.0.6 on an Intel Core i7-10750H CPU. Programming was done in Python, making use of the following libraries: `librosa` [57] and `scipy` for audio processing, `Tensorflow` for deep learning, and `sklearn` for data analysis. The dataset consisted of 16,324 audio samples, which were split into training and test sets using 8:2 ratio. Each audio sample was labeled as either `bird_pest` or `no_pest`. Our experiments primarily sought to demonstrate a systematic method for evaluating machine learning models and optimizing the balance between their performance and size, particularly through the use of KD techniques. Referring to prior KD research [7], we experimented with $T$ values from 1 to 8, and $\alpha$ values from 0.1 to 0.7. Best $T$ value was found to be 7, and best $\alpha$ value was found to be 0.1.

All models were evaluated using standard performance metrics: accuracy (Eq. 5), recall (Eq. 6), precision (Eq. 7), F1 score (Eq. 8), and specificity (Eq. 9):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

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

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

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{8}$$

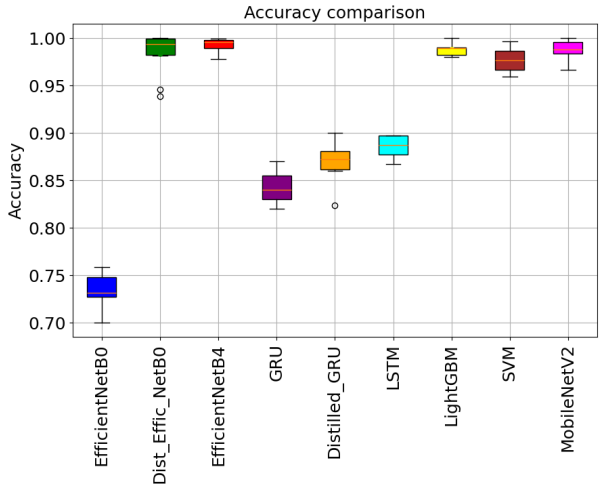
$$\text{Specificity} = \frac{TN}{FP + TN}, \tag{9}$$

where $TP$, $FP$, $TN$, and $FN$ refer to true positives, false positives, true negatives, and false negatives, respectively.

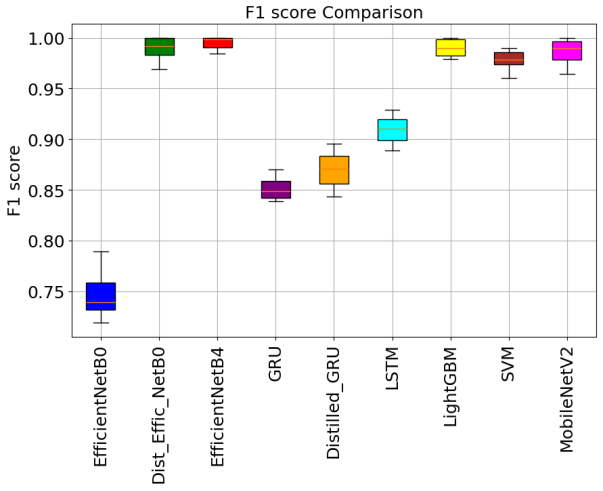
To assess deployment efficiency, we measured two additional implementation metrics: (1) number of parameters $N_p$, i.e., total trainable parameters representing model complexity and memory footprint, and (2) inference time $I_t$, i.e., average time in milliseconds required for a single prediction, measured over 10 consecutive runs. These metrics informed the resource efficiency score, quantifying overall deployment efficiency:

$$\text{Efficiency Score} = \frac{\text{Accuracy} \times 100}{\log_{10}(N_p) \times I_t} \tag{10}$$

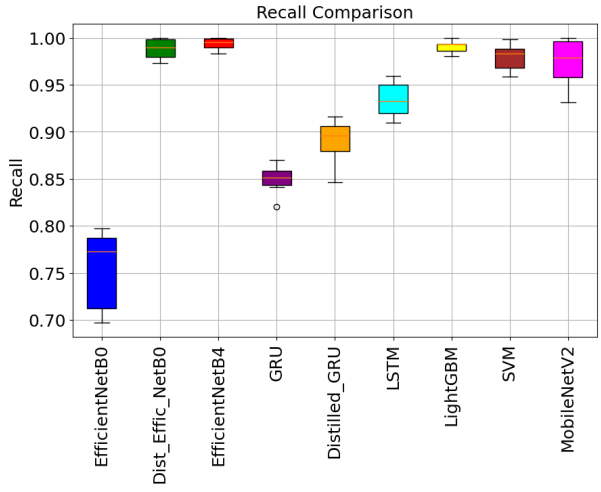
Equation (10) balances accuracy with computational cost, favoring models that achieve high accuracy with minimal parameters and fast inference. We evaluated model compatibility across five representative edge computing platforms with varying computational capabilities: (1) NVIDIA Jetson Nano (quad-core ARM A57 CPU, 128-core Maxwell GPU), (2) Raspberry Pi 4 (quad-core Cortex-A72 CPU), (3) ESP32-S3 (dual-core LX7 microcontroller with neural acceleration), (4) STM32H7 (Cortex-M7 microcontroller), and (5) Arduino Nano 33 BLE (Cortex-M4F microcontroller). These platforms represent a spectrum from



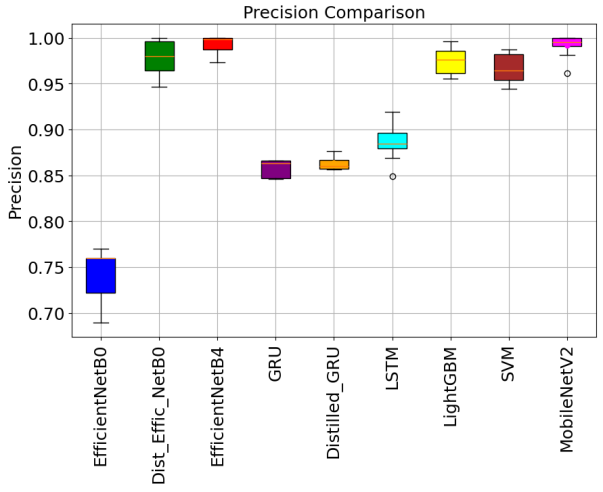

(a) Accuracy comparison



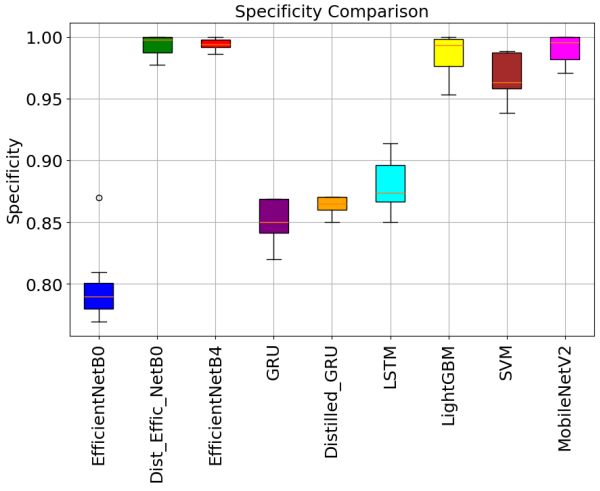

(b) F1 score



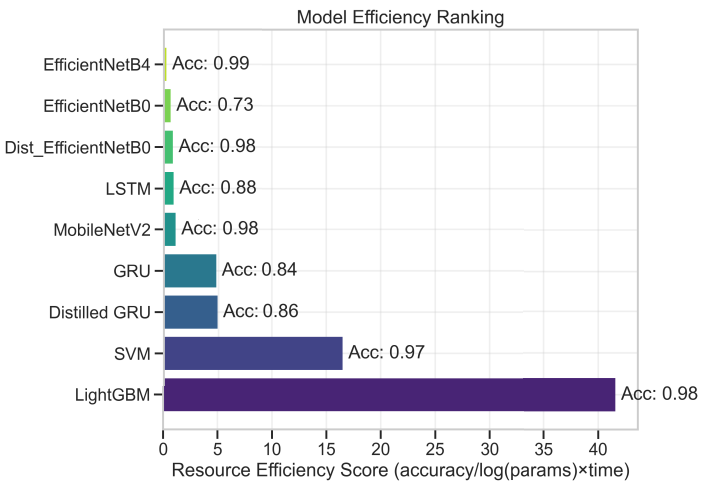

(c) Recall




(d) Precision



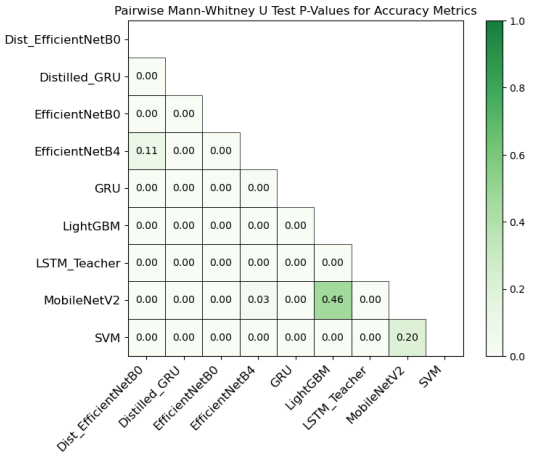

(e) Specificity



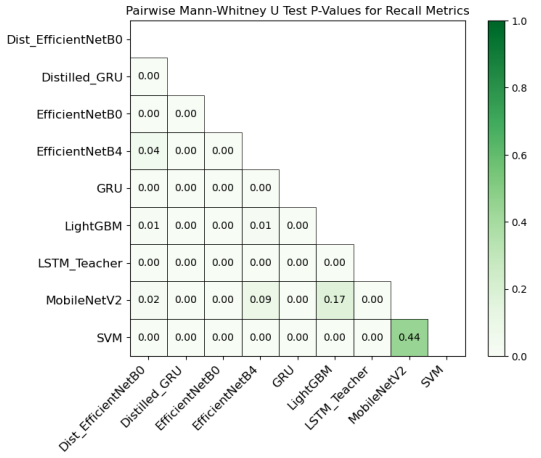

(f) Efficiency score

**FIGURE 8.** **Comparison of accuracy (a), F1 score (b), recall (c), precision (d), and specificity (e) across models using box-and-whisker plots, and efficiency score (f).**

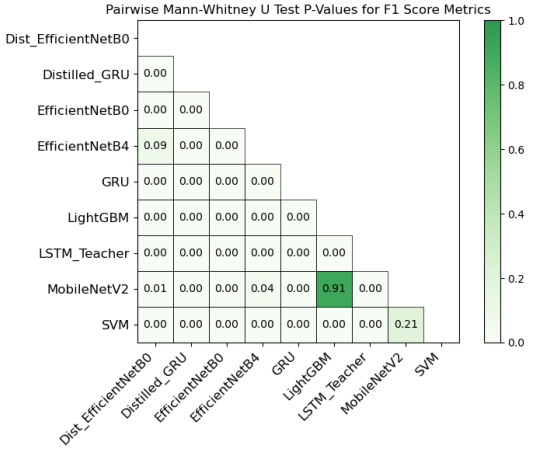
high-performance edge devices to resource-constrained microcontrollers typical in agricultural deployments. The compatibility assessment considered each platform's processing capabilities, memory constraints, and availability



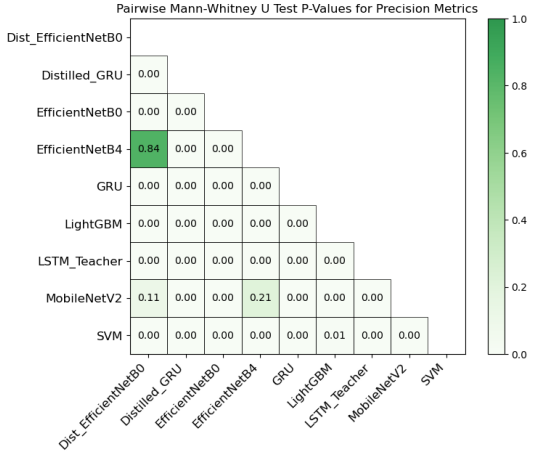

(a) Model performance comparison using p-values based on accuracy metrics.



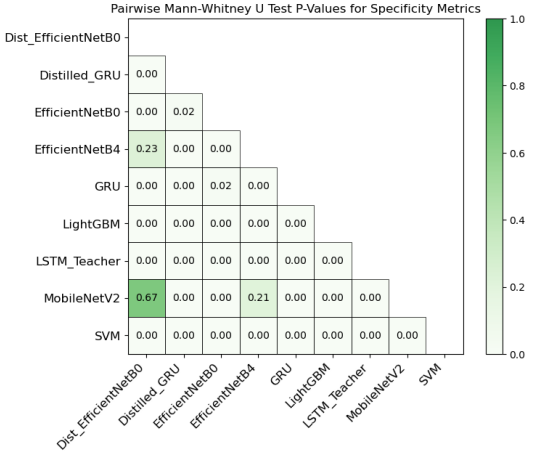

(b) Model performance comparison using p-values based on recall metrics.



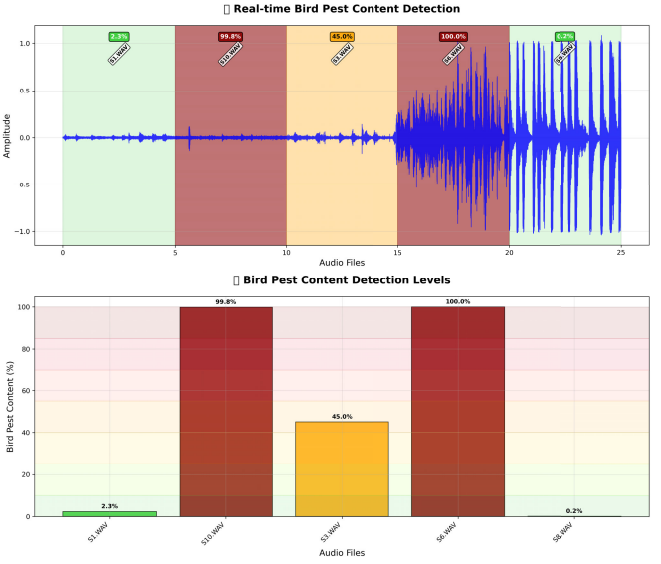

(c) Model performance comparison using p-values based on F1 score metrics.




(d) Model performance comparison using p-values based on precision metrics.




(e) Model performance comparison using p-values based on specificity metrics.




(f) Simulation of LightGBM model predictions on real-world data

**FIGURE 9. Performance comparisons of different models using statistical significance testing (p-values) across multiple metrics (a) to (e) and an example of model behavior while making detections based on the bird pest content detection level (f).**

**TABLE 2.** Performance metrics for ML models (test set). The models with strong performance are highlighted in blue.

| Models | Accuracy | | Recall | | Precision | | F1 score | | Specificity | | Size |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | error | mean | error | mean | error | mean | error | mean | error | |
| Dist_EfficientNetB0 | 0.98 | 0.02 | 0.98 | 0.01 | 0.97 | 0.01 | 0.98 | 0.01 | 0.99 | 0.01 | 22.9 MB |
| EfficientNetB0 | 0.73 | 0.01 | 0.75 | 0.04 | 0.74 | 0.03 | 0.74 | 0.04 | 0.79 | 0.03 | 22.9 MB |
| EfficientNetB4 | 0.99 | 0.01 | 0.99 | 0.00 | 0.99 | 0.01 | 0.99 | 0.00 | 0.99 | 0.00 | 330 MB |
| MobileNetV2 | 0.98 | 0.01 | 0.97 | 0.02 | 0.99 | 0.01 | 0.98 | 0.01 | 0.99 | 0.01 | 14 MB |
| Distilled GRU | 0.86 | 0.02 | 0.89 | 0.02 | 0.86 | 0.00 | 0.87 | 0.01 | 0.86 | 0.00 | 196.8 KB |
| GRU | 0.84 | 0.01 | 0.84 | 0.01 | 0.85 | 0.00 | 0.85 | 0.01 | 0.85 | 0.01 | 196.8 KB |
| LSTM | 0.88 | 0.01 | 0.93 | 0.01 | 0.88 | 0.02 | 0.90 | 0.02 | 0.87 | 0.02 | 29.8 MB |
| LightGBM | 0.98 | 0.00 | 0.99 | 0.00 | 0.97 | 0.01 | 0.98 | 0.00 | 0.98 | 0.01 | 7 KB |
| SVM | 0.97 | 0.01 | 0.97 | 0.01 | 0.96 | 0.01 | 0.97 | 0.00 | 0.96 | 0.01 | 3.3 MB |

**TABLE 3.** Model efficiency metrics with some recommendations for edge device deployment.

| Model | Infer_t_(ms) | Parameters | Eff_Score | Jetson_Nano | Rasp_Pi_4 | ESP32-S3 | STM32H7 | Arduino_Nano_33 |
|---|---|---|---|---|---|---|---|---|
| Dist_EfficientNetB0 | 18 | 4200000 | 0.822 | Excellent | Excellent | Poor | Poor | Not Compatible |
| EfficientNetB0 | 18 | 4200000 | 0.612 | Excellent | Excellent | Poor | Poor | Not Compatible |
| EfficientNetB4 | 58 | 19000000 | 0.235 | Excellent | Excellent | Not Compatible | Not Compatible | Not Compatible |
| MobileNetV2 | 14 | 3500000 | 1.070 | Excellent | Excellent | Poor | Poor | Not Compatible |
| Distilled GRU | 4 | 25000 | 4.889 | Excellent | Excellent | Good | Good | Limited |
| GRU | 4 | 25000 | 4.775 | Excellent | Excellent | Good | Good | Limited |
| LSTM | 15 | 3700000 | 0.893 | Excellent | Excellent | Limited | Limited | Not Compatible |
| LightGBM | 0.6 | 8500 | 41.567 | Excellent | Excellent | Limited | Limited | Not Compatible |
| SVM | 1.2 | 80000 | 16.486 | Excellent | Excellent | Limited | Limited | Not Compatible |

of required libraries and frameworks for model implementation.

## IV. RESULTS AND DISCUSSION

We aimed to systematically compare various ML and deep learning models and evaluate their efficiency to find the optimal solutions. The results are presented in Subsection IV-A. Subsection IV-B presents the statistical hypothesis testing employed to compare the efficacy of the explored models for bird pest detection, as well as an example of the simulation of the model making bird pest sound detection. Subsection IV-C discusses the main observations.

### A. EXPERIMENTAL RESULTS

We evaluated nine models for bird pest detection, as outlined in Section III-C. Table 2 presents performance metrics (accuracy, recall, precision, F1 score, specificity), while Table 3 shows the implementation metrics and the theoretical deployment considerations. EfficientNetB4 achieved highest accuracy (99%) but required substantial resources (19M parameters, 58ms inference). Distilled EfficientNetB0 approached similar performance (98%) with fewer parameters (4.2M). MobileNetV2 also achieved 98% accuracy (3.5M parameters, 14ms inference).

Figure 8(a) to Figure 8(e) visualize metric distributions using box-and-whisker plots. For resource-constrained scenarios, the efficiency score metric (Eq. 10) balanced accuracy against computational requirements. The results of the efficiency score comparison are displayed in Figure 8(f). LightGBM demonstrated the highest efficiency score (41.567) with minimal resources (8.5K parameters, 0.6ms inference) at 98% accuracy. SVM ranked second (16.486) despite larger parameters (80K). Distilled GRU showed good efficiency (4.889) for neural networks (86% accuracy, 25K parameters, 4ms inference). Our projected platform compatibility suggests high-performance edge devices could support all models, while microcontrollers would face limitations. Advanced microcontrollers might support GRU models but show limited compatibility with traditional ML approaches. These findings emphasize considering both performance metrics and implementation requirements when selecting models for resource-constrained agricultural monitoring systems.

### B. HYPOTHESIS TESTING AND EXAMPLE OF DETECTION SIMULATION

We evaluated each model through ten independent runs using five performance metrics: accuracy, F1 score, recall, precision, and specificity. To determine statistical significance

between model performances, we applied the Mann-Whitney U test [58] for pairwise comparisons. Our null hypothesis (H0) stated no difference exists in performance metric distributions between compared models, while the alternative hypothesis (H1) proposed that differences exist. We established a significance threshold of $p < 0.05$ (95% confidence level). When p-values fell below 0.05, we rejected H0, indicating statistically significant performance differences. Pairwise comparison p-values appear in Figure 9(a) through Figure 9(e). Most comparisons revealed significant differences, with two notable exceptions. First, EfficientNetB4 (the overall top performer) and Distilled EfficientNetB0 showed no statistically significant difference, despite Distilled EfficientNetB0 being substantially smaller. Second, LightGBM consistently demonstrated performance statistically equivalent to MobileNetV2. These findings provide compelling evidence that (1) traditional machine learning approaches like LightGBM can match deep learning models in performance and (2) knowledge distillation effectively compresses large models without significant accuracy loss in the agricultural acoustic monitoring context. To demonstrate real-world model performance, we simulated detection on field data using 5 audio samples. Figure 9(f) shows the LightGBM predictions with confidence percentages for bird pest detection. We applied a 50% threshold to classify the presence/absence of bird pests.

### C. DISCUSSION

This research identified significant inverse relationships between computational requirements and detection performance for agricultural bird pest monitoring systems. Traditional ML approaches demonstrated unexpected advantages, with LightGBM achieving 98% of accuracy using only 8,500 parameters and a 0.6 ms inference time, challenging assumptions that deep learning necessarily outperforms traditional methods in resource-constrained contexts. Knowledge distillation proved effective, improving EfficientNetB0 from 73% to 98% accuracy, demonstrating that transfer learning can overcome limitations typically associated with smaller models. Field simulation using LightGBM as an example demonstrated that confidence metrics quantify proportional bird pest acoustic content, with at least 50% of confidence, indicating bird pest dominant recordings. Platform compatibility assessment showed that while high-performance edge devices provide flexibility in model selection, their power requirements and costs may limit widespread adoption. This highlights the importance of application-specific model selection rather than defaulting to the most accurate and/or popular approach.

## V. CONCLUSION

This work established a resource-efficiency framework for acoustic pest detection model selection in resource-limited environments. LightGBM and SVM achieved optimal efficiency-to-accuracy on supported platforms, while distilled GRU provided a neural network alternative, challenging assumptions that effective pest detection requires complex models. While comprehensive field deployment remains future work, this study provides the foundational framework for informed model selection in resource-constrained agricultural settings. Our computational efficiency analysis, combined with emerging prototype development [59], establishes the groundwork for practical deployment systems. Future research priorities include developing adaptive multinomial classification systems capable of distinguishing between various pest species, beneficial birds, and environmental sounds across diverse agricultural environments, recognizing that species composition varies by geographic region and farm type. Additional priorities include targeted model compression, power optimization, and evaluation of deployment considerations, including weather durability. These advances will enable affordable detection systems that adapt to local ecosystems while protecting crops and preserving beneficial species.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the support of PASET-RSIF scholarships awarded to Micheline Kazeneza and Destiny Kwabla Amenyedzi through the University of Rwanda's African Centre of Excellence in the Internet of Things (ACEIoT).

## REFERENCES

[1] M. A. Kale, N. Dudhe, R. Kasambe, and P. Bhattacharya, "Crop depredation by birds in deccan Plateau, India," *Int. J. Biodiversity*, vol. 2014, pp. 1–8, Sep. 2014.

[2] M. Issa and M. El-Bakhshawngi, "An estimation of bird damages on some field, vegetable and fruit crops at sharkia governorate, Egypt," *Zagazig J. Agricult. Res.*, vol. 45, no. 4, pp. 1273–1281, Jul. 2018.

[3] P. Pukhamwong and C. Boonyasiriwat, "An implementation of a recurrent neural network for 1D acoustic waveform inversion," *J. Phys., Conf. Ser.*, vol. 1719, no. 1, Jan. 2021, Art. no. 012035.

[4] L. Wyse, "Audio spectrogram representations for processing with convolutional neural networks," 2017, *arXiv:1706.09559*.

[5] A. D. P. Ramirez, J. I. de la Rosa Vargas, R. R. Valdez, and A. Becerra, "A comparative between mel frequency cepstral coefficients (MFCC) and inverse mel frequency cepstral coefficients (IMFCC) features for an automatic bird species recognition system," in *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, Nov. 2018, pp. 1–4.

[6] M. Kazeneza, D. K. Amenyedzi, A. Vodacek, D. Hanyurwimfura, and E. Ndashimye, "Bird sound classification using GLCM features and LightGBM applied to farm monitoring," in *Proc. 11th Int. Conf. Intell. Comput. Wireless Opt. Commun. (ICWOC)*, Jun. 2023, pp. 20–24.

[7] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[8] F. Sarfraz, E. Arani, and B. Zonooz, "Knowledge distillation beyond model compression," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 6136–6143.

[9] D. Bagchi and W. Hartmann, "Learning from the best: A teacher–student multilingual framework for low-resource languages," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6051–6055.

[10] G. Mahjoub, M. K. Hinders, and J. P. Swaddle, "Using a 'sonic net' to deter pest bird species: Excluding European starlings from food sources by disrupting their acoustic communication," *Wildlife Soc. Bull.*, vol. 39, no. 2, pp. 326–333, Jun. 2015.

[11] T. Scott Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," *Bird Conservation Int.*, vol. 18, no. S1, pp. S163–S173, Sep. 2008.

[12] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," *Ecol. Informat.*, vol. 61, Mar. 2021, Art. no. 101236. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954121000273
[13] D. Stowell, M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin, "Automatic acoustic detection of birds through deep learning: The first bird audio detection challenge," *Methods Ecol. Evol.*, vol. 10, no. 3, pp. 368–380, Mar. 2019.
[14] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[15] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2019, pp. 6105–6114.
[16] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.
[17] J. Kang, W.-Q. Zhang, and J. Liu, "Gated recurrent units based hybrid acoustic models for robust speech recognition," in *Proc. 10th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Oct. 2016, pp. 1–5.
[18] Y. Zhao, J. Li, S. Xu, and B. Xu, "Investigating gated recurrent neural networks for acoustic modeling," in *Proc. 10th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Oct. 2016, pp. 1–5.
[19] Q. Lu, Y. Li, Z. Qin, X. Liu, and Y. Xie, "Speech recognition using EfficientNet," in *Proc. 5th Int. Conf. Multimedia Syst. Signal Process.*, May 2020, pp. 64–68.
[20] D. T. Speckhard, K. Misiunas, S. Perel, T. Zhu, S. Carlile, and M. Slaney, "Neural architecture search for energy-efficient always-on audio machine learning," *Neural Comput. Appl.*, vol. 35, no. 16, pp. 12133–12144, Jun. 2023.
[21] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Dec. 2017, pp. 3146–3154.
[22] C. Cortes, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[23] C. Zieger and M. Omologo, "Acoustic event classification using a distributed microphone network with a GMM/SVM combined algorithm," in *Proc. Interspeech*, Sep. 2008, pp. 115–118.
[24] F. König, C. Sous, A. O. Chaib, and G. Jacobs, "Machine learning based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems," *Tribol. Int.*, vol. 155, Dec. 2020, Art. no. 106811.
[25] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 6, pp. 1789–1819, Mar. 2021.
[26] A. P. Hill, P. Prince, J. L. Snaddon, C. P. Doncaster, and A. Rogers, "AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment," *HardwareX*, vol. 6, Oct. 2019, Art. no. e00073. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468067219300306
[27] R. Manzano-Rubio, G. Bota, L. Brotons, E. Soto-Largo, and C. Pérez-Granados, "Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species," *Ecol. Informat.*, vol. 72, Dec. 2022, Art. no. 101910.
[28] J. Marchal, F. Fabianek, and Y. Aubry, "Software performance for the automated identification of bird vocalisations: The case of two closely related species," *Bioacoustics*, vol. 31, no. 4, pp. 397–413, Jul. 2022.
[29] A. A. Akinrinmade, E. Adetiba, J. A. Badejo, and S. I. Popoola, "Effect of spectrogram preprocessing and enhancement on speaker recognition performance," in *Proc. Int. Conf. Sci., Eng. Bus. Sustain. Develop. Goals (SEB-SDG)*, vol. 1, Apr. 2023, pp. 1–6.
[30] O. Özhan, "Short-time-Fourier transform," in *Basic Transforms for Electrical Engineering*. Cham, Switzerland: Springer, 2022, pp. 441–464.
[31] M. Dörfler, R. Bammer, and T. Grill, "Inside the spectrogram: Convolutional neural networks in audio processing," in *Proc. Int. Conf. Sampling Theory Appl. (SampTA)*, Jul. 2017, pp. 152–155.
[32] M. Küçükakarsu, A. Kavsaoğlu, F. Alenezi, A. Alhudhaif, R. Alwadie, and K. Polat, "A novel automatic audiometric system design based on machine learning methods using the brain's electrical activity signals," *Diagnostics*, vol. 13, no. 3, p. 575, Feb. 2023.
[33] A. Eledkawy, T. Hamza, and S. El-Metwally, "Precision cancer classification using liquid biopsy and advanced machine learning techniques," *Sci. Rep.*, vol. 14, no. 1, p. 5841, Mar. 2024.
[34] S. Rahman, M. Hasan, and A. K. Sarkar, "Prediction of brain stroke using machine learning algorithms and deep neural network techniques," *Eur. J. Electr. Eng. Comput. Sci.*, vol. 7, no. 1, pp. 23–30, Jan. 2023.
[35] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "Machine learning techniques for classifying dangerous asteroids," *MethodsX*, vol. 11, Dec. 2023, Art. no. 102337.
[36] D. Tomar and S. Agarwal, "A comparison on multi-class classification methods based on least squares twin support vector machine," *Knowl.-Based Syst.*, vol. 81, pp. 131–147, Jun. 2015.
[37] A. Rabaoui, M. Davy, S. Rossignol, and N. Ellouze, "Using one-class SVMs and wavelets for audio surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 4, pp. 763–775, Dec. 2008.
[38] J.-C. Wang, L.-X. Lian, Y.-Y. Lin, and J.-H. Zhao, "VLSI design for SVM-based speaker verification system," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 7, pp. 1355–1359, Jul. 2015.
[39] G. Muhammad, Y. A. Alotaibi, M. Alsulaiman, and M. N. Huda, "Environment recognition using selected mpeg-7 audio features and mel-frequency cepstral coefficients," in *Proc. Int. Conf. Digit. Image Process. Pattern Recognit.*, Jun. 2019, pp. 120–125.
[40] U. Kumar, U. Thakker, P. Hille, and A. Lauber, "Deploying machine learning algorithms on resource-constrained IoT devices: A review," *Sensors*, vol. 22, no. 4, p. 1550, 2022.
[41] K. He, R. Girshick, and P. Dollár, "Rethinking ImageNet pre-training," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4917–4926.
[42] V.-T. Hoang and K.-H. Jo, "Practical analysis on architecture of EfficientNet," in *Proc. 14th Int. Conf. Human Syst. Interact. (HSI)*, Gdansk, Poland, Jul. 2021, pp. 1–4.
[43] Y. Gong, Y.-A. Chung, and J. Glass, "AST: Audio spectrogram transformer," in *Proc. Interspeech*, Aug. 2021, pp. 571–575.
[44] K. Dong, C. Zhou, Y. Ruan, and Y. Li, "MobileNetV2 model for image classification," in *Proc. 2nd Int. Conf. Inf. Technol. Comput. Appl. (ITCA)*, Dec. 2020, pp. 476–480.
[45] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[46] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Trans. Audio, Speech, Languages Process.*, vol. 28, pp. 2880–2894, 2020.
[47] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao, J. Wu, L. Zhou, S. Ren, Y. Qian, Y. Qian, J. Wu, M. Zeng, X. Yu, and F. Wei, "WavLM: Large-scale self-supervised pre-training for full stack speech processing," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 6, pp. 1505–1518, Oct. 2022.
[48] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
[49] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "Wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2020, pp. 12449–12460.
[50] A. Blattmann, R. Rombach, K. Oktay, J. Müller, and B. Ommer, "Semi-parametric neural image synthesis," 2022, *arXiv:2204.11824*.
[51] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
[52] S. Kahl, A. K. Navine, T. Denton, H. Klinck, P. J. Hart, H. Glotin, H. Goëau, W.-P. Vellinga, R. Planqué, and A. Joly, "Overview of BirdCLEF 2022: Endangered bird species recognition in soundscape recordings," in *Proc. Work. Notes Conf. Labs Eval. Forum (CLEF)*, vol. 3180, Bologna, Italy, Sep. 2022, pp. 1929–1939. [Online]. Available: https://ceur-ws.org/Vol-3180/paper-154.pdf
[53] A. Alkhulaifi, F. Alsahli, and I. Ahmad, "Knowledge distillation in deep learning and its applications," *PeerJ Comput. Sci.*, vol. 7, p. e474, Apr. 2021.
[54] Q. Huang, X. Wu, Q. Wang, X. Dong, Y. Qin, X. Wu, Y. Gao, and G. Hao, "Knowledge distillation facilitates the lightweight and efficient plant diseases detection model," *Plant Phenomics*, vol. 5, p. 62, Jul. 2023.
[55] Y. Wang, H. Li, L.-P. Chau, and A. C. Kot, "Embracing the dark knowledge: Domain generalization using regularized knowledge distillation," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 2595–2604.

[56] M. Yuan, B. Lang, and F. Quan, "Student-friendly knowledge distillation," *Knowl.-Based Syst.*, vol. 296, Jul. 2024, Art. no. 111915.
[57] S. Suman, K. S. Sahoo, C. Das, N. Z. Jhanjhi, and A. Mitra, "Visualization of audio files using librosa," in *Proc. 2nd Int. Conf. Math. Model. Comput. Sci. (ICMMCS)*. Cham, Switzerland: Springer, Jan. 2022, pp. 409–418.
[58] N. Nachar, "The mann-whitney U: A test for assessing whether two independent samples come from the same distribution," *Tuts. Quant. Methods Psychol.*, vol. 4, no. 1, pp. 13–20, Mar. 2008.
[59] D. K. Amenyedzi, M. Kazeneza, I. I. Mwaisekwa, F. Nzanywayingoma, P. Nsengiyumva, P. Bamurigire, E. Ndashimye, and A. Vodacek, "System design for a prototype acoustic network to deter avian pests in agriculture fields," *Agricult.*, vol. 15, no. 1, p. 10, Dec. 2024.

**MICHELINE KAZENEZA** (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science and technology from Saint Petersburg Electrotechnical University, Saint Petersburg, Russia, in 2012 and 2014, respectively. Currently, she is pursuing the Ph.D. degree with the ACEIoT, University of Rwanda. In 2024, she completed an Internship from the University of Pretoria through the International Partner Institutions Program. From 2015 to 2021, she was an Assistant Lecturer with the University of Burundi. Her research interests include sound detection, machine learning, model optimization, and the Internet of Things.

**ANNA SERGEEVNA BOSMAN** (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Pretoria, South Africa, in 2012 and 2019, respectively. She is currently appointed as a Senior Lecturer with the Department of Computer Science, University of Pretoria. Her research interests include deep neural networks, loss landscape analysis, knowledge distillation, energy-efficient models, meta-learning, and computer vision. She has been a member of the IEEE Computational Intelligence Society (CIS), since 2014, and serves on the IEEE CIS Neural Networks Technical Committee (NNTC).

**DESTINY KWABLA AMENYEDZI** (Student Member, IEEE) received the B.E.D. degree in information technology from the University of Education, Winneba, in 2009, and the M.Phil. degree from the Kwame Nkrumah University of Science and Technology, in 2014. Currently, he is pursuing the Ph.D. degree with the ACEIoT, University of Rwanda. His professional experience includes teaching with Kopeyia Bloomfield Basic School, from 2002 to 2005, and training teachers with the St. Francis College of Education, from 2009 to 2021. With six peer-reviewed publications, his research focuses on acoustic monitoring for agriculture, mobile learning applications, and remote sensing.

**DAMIEN HANYURWIMFURA** received the B.Sc. degree in computer engineering and information technology from the University of Rwanda (former KIST), in 2005, and the M.Sc. and Ph.D. degrees in computer science and technology from Hunan University, China, in 2010 and 2015, respectively. He is currently a Professor of computer and software engineering and the Acting Director of the African Center of Excellence in Internet of Things (ACEIoT), College of Science and Technology, University of Rwanda. He was the Head of the Ph.D. Studies and Research, ACEIoT, for four years. He has published and co-authored more than 65 research papers in leading international journals and conferences. His research interests include the Internet of Things, artificial intelligence, data mining, and machine learning.

**EMMANUEL NDASHIMYE** (Member, IEEE) received the B.Sc. degree from the University of Rwanda, in 2005, the M.Sc. degree (Hons.) from Huazhong University of Science and Technology, China, in 2009, and the Ph.D. degree from Auckland University of Technology, New Zealand, in 2018. He is currently an Assistant Teaching Professor with Carnegie Mellon University Africa. His research interests include vehicle communications, networking technologies for mobile nodes, 5G networks, network security, and the IoT. He formerly served as an Editor for *Engineering Science and Technology*, from 2019 to 2022.

**ANTHONY VODACEK** (Senior Member, IEEE) received the B.S. degree from the University of Wisconsin–Madison, in 1981, and the M.S. and Ph.D. degrees from Cornell University, in 1985 and 1990, respectively. He is currently a Full Professor of imaging science with Rochester Institute of Technology. He supports ACEIoT through a Ph.D. Student Supervision and was a Fulbright Specialist. He supports African Great Lakes monitoring initiatives for the African Center for Aquatic Research and Education (ACARE). His research interests include multi-modal remote sensing for terrestrial and aquatic systems. He serves as the IEEE-GRSS Global Activities Liaison to Sub-Saharan Africa and the Scientific Advisor to Space4Innovation. He is an Associate Editor for *Journal of Great Lakes Research*.