

APPLIED RESEARCH

Event Detection and Classification for Long Range Sensing of Elephants Using Seismic Signals

JALIYA L. WIJAYARAJA¹, (Member, IEEE), JANAKA L. WIJEKOON^{2,3}, (Senior Member, IEEE), AND MALITHA WIJESUNDARA¹, (Member, IEEE)

¹Sri Lanka Institute of Information Technology, Malabe 10115, Sri Lanka

²Victorian Institute of Technology, Adelaide Campus, Adelaide, SA 5000, Australia

³Department of System Design Engineering, Keio University, Yokohama 108-0073, Japan

Corresponding author: Jaliya L. Wijayaraja (jaliya.wijayaraja@gmail.com)

This work was supported by the Sri Lanka Institute of Information Technology (SLIIT).

ABSTRACT Detecting elephants through seismic signals is an emerging research topic aimed at developing solutions for Human-Elephant Conflict (HEC). Despite the promising results, such solutions heavily rely on manual classification of elephant footfalls, which limits their applicability for real-time classification in natural settings. To address this limitation and build on our previous work, this study introduces a classification framework targeting resource-constrained implementations, prioritizing both accuracy and computational efficiency. As part of this framework, a novel event detection technique named Contextually Customized Windowing (CCW), tailored specifically for detecting elephant footfalls, was introduced, and evaluations were conducted by comparing it with the Short-Term Average/Long-Term Average (STA/LTA) method. The yielded results show that the maximum validated detection range was 155.6 m in controlled conditions and 140 m in natural environments. Elephant footfall classification using Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel demonstrated superior performance across multiple settings, achieving an accuracy of 99% in controlled environments, 73% in natural elephant habitats, and 70% in HEC-prone human habitats, the most challenging scenario. Furthermore, feature impact analysis using explainable AI identified the number of Zero Crossings and Dynamic Time Warping (DTW) Alignment Cost as the most influential factors in all experiments, while Predominant Frequency exhibited significant influence in controlled settings.

INDEX TERMS Elephant detection, human–elephant conflict, seismic signals, ANN, SVM, ExAI, STA/LTA, embedded system.

I. INTRODUCTION

The human-elephant conflict (HEC) poses significant challenges in wildlife management, socio-economic stability, and environmental conservation across a range of countries [1]. Notably, Sri Lanka faces acute challenges from HEC due to its high density of elephant populations [2], [3], reporting the highest annual elephant fatalities and the second-highest human casualties globally [4]. A promising approach to managing HEC involves the early detection of elephants using non-invasive methods [5]. Among these, seismic signal-based detection demonstrates a significant potential [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos¹.

Seismic signals generated by elephants propagate over long distances with minimal interference, [5], [7], [8], while retaining distinctive characteristics that enable their differentiation from other environmental seismic disturbances [5], [6]. Notably, seismic-based detection offers distinct advantages over other non-invasive methods, such as visual, [9], [10], [11], [12], [13], [14], [15], [16], [17], [18] acoustic, [19], [20], [21], [22], [23], [24], [25], [26], [27] and infrasonic, [21], [28] detection. Specifically, seismic detection has demonstrated strong potential for reliable, long-range identification of elephants, with robust and sustainable performance under tropical conditions [5]. Therefore, we previously developed an embedded instrumentation system featuring a novel geophone-sensor interface,

specifically tailored for detecting elephant-generated seismic signals over extended distances [5].

The introduced instrumentation system in [5] demonstrate a capability to detect elephants at a maximum distance of 155.6 m with an overall accuracy of 99.5%. This system was specifically focused on Sri Lankan elephants (*Elephas maximus maximus*) a distinct subspecies of Asian elephants [29], [30], and was tested under controlled, partially-controlled, and natural environments with minimal noise. However, the classification process relied heavily on manual intervention. Researchers manually inspected each recording, compared with reference videos, and split and segmented the footfall events for further analysis. Subsequently, these events were categorized into classes (elephant footfall, human footfall, noise from a motorcycle) based on their predominant frequency, utilizing a Decision Tree (DT) model.

While manual splitting is effective for research applications, developing a field-implementable early detection solution for HEC necessitates automating the entire process from signal acquisition to classification. Hence, automating the process by utilizing event detection and extraction algorithms becomes imperative. Given the natural environments often prone to significant ambient noise, the existing method in [5], which relies on a single feature for event detection, may fail in complex natural environments despite showing satisfactory results in controlled settings. Therefore, it is vital to explore additional features and develop more sophisticated classification models tailored to Sri Lankan elephants and resilient to the natural environmental conditions prevalent in Sri Lanka.

To this end, this study first introduces a dedicated filter designed to suppress seismic noise commonly present in natural environments. Subsequently, a novel event detection method termed the context-customized window (CCW), which is specifically tailored for elephant footfall signals, is proposed. The CCW is then compared with two other existing event detection methods, namely, Short-Term Average/Long-Term Average (STA/LTA) and the modified Energy Ratio (MER), which are commonly used in seismology applications [31]. These event detection methods will be evaluated based on their accuracy and execution time. Additionally, this study discusses nine distinct features, categorized into temporal, spectral, pattern-matching, and statistical types, that were extracted from the detected events for further classification.

This study assesses two classification models for their ability to classify the extracted features and identify signals generated by elephant footfalls. A Support Vector Machine (SVM) was employed as a low-complexity model suitable for resource-constrained environments [32], [33], such as embedded systems. Same datasets were also used to train an Artificial Neural Network (ANN) [32] as a reference model. Both models were trained and tested using datasets from various conditions, including controlled environments and natural settings resembling HEC-prone areas. Finally, this study analyzes the impact of each feature on the

classification accuracy of elephants in both natural and controlled conditions to guide future improvements.

II. BACKGROUND SURVEY

Existing studies on event detection and classification using seismic signals have focused mainly on specific domains [5], [34], [35], [36]. Notably, studies such as [36], [37], [38], and [39] present successful approaches utilizing seismic signal detection and classification in recognizing human activities, confirming that seismic signals are a viable means of footfall classification and identification. Among the aforementioned studies, [38] proposed a strategy specifically tailored for embedded systems, integrating event detection, feature extraction, and event classification.

The study in [38] highlights the importance of event detection, particularly when event detection is employed to identify the presence and location of significant seismic events within a signal. Specifically, [38] introduces event detection methods including amplitude thresholding, kurtosis-based event detection, STA/LTA, and MER which are utilized before classification. Similarly, [40] proposed an advanced adaptive approach for event detection using STA/LTA method. However, similar to [38], the scope of [40] also remains limited to human footfall.

Nevertheless, the study [41], which is closely aligned with one of the objectives of this study, introduced three methods to detect elephant footfall events: the STA/LTA algorithm, continuous wavelet transforms (CWT), and amplitude spectrum of the Fourier transform (ASFT). However, [41] is limited to event detection; it does not extend to classifying the detected events to differentiate elephants' seismic signals from other seismic activities.

Conversely, [42] employed kurtosis analysis to distinguish seismic events from ambient noise, followed by classification that achieved 89% accuracy in identifying elephants using feature thresholding based on the standard deviation of the signal, frequency peak, spectral centroid, and frequency distribution. Despite demonstrating the potential importance of event detection methods, the study detected only 54% of footsteps and was limited to a range of 8 to 30 meters, focusing on African elephants.

Despite the lack of event detection, an essential component for achieving practical HEC solutions, several studies successfully concluded that various distinctive characteristics of seismic waves generated by elephants can be used to classify and differentiate them from seismic signals generated by other animals or human activities [6], [8]. Such studies are summarized in Table 1.

As detailed in the Table 1, an early study conducted in 2005 by Wood et al. [6] employed a discriminant analysis to achieve a maximum accuracy of 82% in detecting seismic signals generated by elephant footfalls, with a relatively higher detection range of approximately 100 meters for African elephants in their natural habitats. This study used a sound card to collect seismic signals, and a correlation coefficient was used for event classification. Later, the

TABLE 1. Summary of literature on classification and detection of elephants using seismic signals.

Ref.	Acquisition Method	Elephant Type	Event Type	Detection Range	Classification Method	Features Employed	Reported Accuracy
[6]	So	African	L	~100 m	Discriminate analysis	Correlation coefficient of power spectra	82%
[43]	So	Asian	L	CL	MLP	Signal recordings	50%
[8]	Se	African	V/L	150 m	CNN	Spectrogram	73-90%
[44]	Em	Asian	L	3-5 m	ANN	Signal recordings	100%
[45]	Em	NA	L	NA	Logistic regression	Signal recordings	93%
[46]	Em	Asian	L	NA	SVM	Ref Table 2	73%
[47]	Em	Asian	V	NA	Ridge classifier**	MFCC**	97.05%
[48]	Em	Asian	L	1-25 m	CNN**	Signal recordings	98.03%
[49]	Em	Asian	L	2 m	Feature thresholding	Signal frequency and amplitude, Cumulative volume of vibration	80%
[50]	Em	Asian	L	NA	CNN-BiLSTM**	29 multi-domain features ¹	77.89%
[51]	Se	African	L	150 m	Support vector classifier	Scattering transform features extracted from seismogram	77%
[52]	Em	NA	L	NA	Amplitude thresholding	Average signal amplitude	NA

Note: So – Sound card; Se – Seismometers; Em – Embedded Systems; L – Locomotion; V – Vocalization; CL – Close to path; MLP – Multi Layer Perceptron; CNN – Convolutional Neural Network; ANN – Artificial Neural Network; MFCC – Mel Frequency Cepstral Coefficients; SVM – Support Vector Machine; SVC – Support Vector Classifier; LSTM – Long Short-Term Memory. ** This study trialed multiple methods; the given method is the best among them.

TABLE 2. Features utilized in related studies for identifying footfall seismic signals of elephants.

Feature Type	Feature
Temporal	Energy of Signal, Zero Crossing Rate, Skewness, Kurtosis, Shannon Entropy
Spectral	Energy of Frequency Bands, Spectral Cepstral
Cepstral	Log energy of MFCC

same setup was followed by [43] for Sri Lankan elephants employing the multilayer perception (MPL) method that directly used the signal recordings. Although the experiments were conducted in a controlled environment in a very close area, this study reported the lowest accuracy among comparable studies. According to the authors, the use of a generic sound card instead of a high-end sound card (that is, the Digigram vx pocket v2 used in [6] hindered the performance. However, it is a notable fact that reliance on sound cards — a computer peripheral designed for indoor use, obviously limits the feasibility as field-deployable solutions for HEC due to limited durability in outdoor tropical terrains.

In a similar study, [8] employed standard seismic data acquisition methods (i.e., 22 6TD Guralp seismometers) to record the seismic signal of African elephants in their natural habitats. This study achieved 80–90% accuracy in classifying elephant footfalls at a range of 100 m by using spectrograms of elephant footfall recordings together with a Convolutional Neural Network (CNN). However, accuracy dropped to 73% when the range was increased to 150 m, highlighting the critical impact of range on the detection accuracy of the classifier. Furthermore, the study highlighted that the higher detection ranges were obtained under the limitation of using trained and tested data from the same seismometer and from matching environmental conditions from different locations, which restricts the classifier's usability as an applicable solution for HEC. Using the same dataset, [51] improved

the accuracy to 77% at 150 meters by employing a Support Vector Classifier (SVC). Nevertheless, the dependence on high-cost seismometers limits their feasibility and scalability of a cost-effective solution for mitigating HEC.

The recent studies highlight the fact that the generic embedded systems are also emerging as a feasible approach for acquiring seismic signals and classifying elephants, offering customizability as a practical solution for HEC [5]. Notably, authors from 'Gaja-Mithuru' reported that 100% of classification accuracy was achieved by using an artificial neural network, with sensors deployed at close distances of 1–3 m [44]. However, the model in [44] was trained to distinguish and categorize walking and running, and no effort was made to distinguish the elephants' footfalls from other ambient signals.

Similarly, [48] reported the highest accuracy of 98.03% using their own CNN model named 'GajGamini' while comparing it with several other classification methods including Random Forest, Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, K-Nearest Neighbors (KNN), DT, MLP, and a CNN model introduced in [8] named 'SeisSavanna.' Both these methods used signal recordings without any feature extraction for the classification. While conventional machine learning models typically exhibit lower computational complexity [38], this study in [48] highlights that, in real-time implementation, the proposed CNN model outperformed conventional machine learning models by eliminating the need for explicit feature extraction. This, in turn, underscores the importance of optimizing conventional machine learning methods through

¹Features are not explicitly mentioned.

the elimination of less significant features to maintain computational efficiency.

Additionally, [45] achieved a 93% accuracy of identifying elephants using logistic regression directly on the signal recordings. Parihar et al. [46], on the other hand, suggested using nine different features (presented in Table 2) to classify the elephants instead of relying on raw signal recordings. These features represent multiple domains of the signal, including temporal, spectral, and cepstral characteristics. Subsequently, these features were used for the classification of seismic signals using an SVM classifier. However, the results revealed that the classification achieved only 73% accuracy under natural environments. Building on the same dataset, [50] improved the accuracy by approximately 5% by employing a CNN-BiLSTM model with 29 multi-domain features.

Vidunath et al. in [47], in a recent study, used vocalization for elephant classification instead of using footfalls. They employed a combination of feature extraction methods to extract spectral features, along with various classifiers. Specifically, Mel-frequency cepstral coefficients (MFCC) were classified with a Ridge Classifier, SVM, and Logistic Regression; Hjorth parameters were classified using a Decision Tree (DT), AdaBoost Classifier, and Random Forest Classifier; and spectral energy distribution was classified with a Light Gradient Boosting Machine (LGBM), Gradient Boosting Classifier, and AdaBoost Classifier. The best performance achieved was 97.05% with MFCC and a Ridge Classifier. Moreover, another publication [53] from the same study stated that only seven rumbles were recorded during several hours of observation. Additionally, vocalizations were found to propagate over shorter distances compared to footfall signals [54], highlighting the reduced potential of vocalizations for long-range and accurate classification compared to locomotion.

Interestingly, recent studies in this domain suggest more integrated approaches that combine real-time seismic signal classification with remedial interventions [49], [52]. Both studies relied on basic threshold-based detection techniques, and [49] did not focus on field-level detection accuracy or operational range, whereas [49] employed a two-sensor string-based configuration instead of conventional geophones, reporting 80% accuracy with a limited detection range of 2 meters. Most importantly, [49] also identifies critical limitations, including misclassification due to overlapping frequency patterns from non-target animal groups and performance degradation caused by topographical conditions. These findings highlight the necessity for more advanced classification methodologies to enable accurate, long-range detection in diverse environmental contexts.

Apart from our previous work in [5], the handful of studies conducted using generic embedded systems [44], [45], [46], [47], [48] showed a very limited range in detecting elephant footfalls. Although the reported accuracies are satisfactory, most of these studies derived their results from data collected in controlled environments, as mentioned in [44], [46],

[47], and [48], hindering the implementations in natural HEC prone habitats. Additionally, the study in [45] utilized secondary data for which the species of the elephants under study and the nature of the environment were not confirmed. Still, related comprehensive studies have concluded that the environmental conditions of the training dataset are critical to the accuracy of the final implementation [8]. Therefore, it is essential to conduct further experiments to train and analyze the performance of the classifiers in natural habitats of elephants where HEC is prone to occur, particularly for Sri Lankan elephants.

In summary, there is significant potential to automate the process outlined in our previous work [5] by utilizing event detection, event extraction, and feature extraction as presented in related studies. Also, the available literature reveals that there is a gap in developing a classifier that is resilient to natural environments prone to HEC, particularly for Sri Lankan elephants, while enabling long-range detection. Moreover, existing literature has concluded that feature extraction is computationally expensive for classifying and identifying elephants using seismic signals [48]. Thus, further analysis is required to identify the most impactful features to optimize conventional machine learning models, particularly for resource-constrained environments such as embedded systems.

III. METHODOLOGY

This study introduces a systematic process for classifying and identifying elephant footfalls, comprising event detection, event extraction, feature extraction, and classification. Subsequently, explainable AI (XAI) techniques were employed to analyze the contribution of each feature to classification accuracy. This process, illustrated as a pipeline in Fig. 1, is discussed in detail in the subsequent sections.

A. EVENT DETECTION

Accurate detection of elephant footfall events, while effectively suppressing seismic noise, is essential for the long-term continuous operation of monitoring systems. The use of computationally intensive algorithms for feature extraction and classification often results in inefficient resource utilization, particularly in embedded systems with limited computational capabilities. Our previous study [5] highlighted instances where the instrumentation system generated low-amplitude noise patterns without meaningful events, typically occurring in the absence of elephant or other significant activity within the monitored area. Furthermore, as noted by [46], even in the presence of elephants, their natural behavior characterized by intermittent walking interspersed with pauses for feeding or social interactions, results in substantial temporal gaps between footfall events. Similar behavioral patterns were observed in our field studies conducted in Digampathaha, Sri Lanka [5]. Consequently, this study prioritizes the initial detection and extraction of potential elephant footfall events from seismic signals, while minimizing the noise.

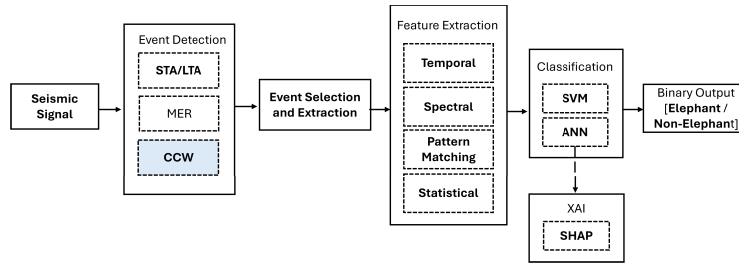


FIGURE 1. Overview of the proposed seismic signal classification pipeline.

In this study, three distinct event detection methods were implemented for comparison and benchmarking: STA/LTA, MER, and CCW. Among these, CCW is a novel approach introduced specifically in this study, while STA/LTA and MER are well-established techniques commonly used in seismological research [31], [55]. To establish a baseline for event detection, each scenario was visually inspected using synchronized video recordings from field studies, with relevant events manually identified and annotated. This manually annotated dataset is considered as the ground truth for the evaluations.

Given the importance of accurately defining the duration of seismic events for precise event detection, this study manually and systematically selected 50 samples of elephant footstep events based on waveform morphology from seismic signal recordings obtained during field studies.² The average duration of the seismic signals generated by a single elephant footfall was then calculated and designated as the reference length of a seismic event for Sri Lankan elephants, denoted as L_{Elp} . The minimum and maximum marginal durations of these samples were identified and labeled as L_{Elp_Min} and L_{Elp_Max} , respectively.

1) SHORT-TERM AVERAGE/LONG-TERM AVERAGE (STA/LTA)

The STA/LTA method calculates the ratio of average energy in a short-time window to that in a long-time window [38], [56]. The short-time window precedes the long trailing window. According to [38], the energy ratio is calculated using (1) where $er[n]$ represents the energy ratio at testing index n . s and l denote the length of the short and long-time windows. Here, x represents the signal function.

$$er[n] = \frac{\frac{1}{s} \sum_{n_1=l}^{l+s} (x[n_1])^2}{\frac{1}{l} \sum_{n_2=n}^{n+l} (x[n_2])^2} \quad (1)$$

According to general guidelines, the length of the short-time window is typically chosen to be shorter than the shortest expected event [56]. Additionally, it is recommended that the length of the long time window be 5 to 10 times that of the short time window [38], [57]. Consequently, a ratio of 5 was approximated based on observations of the elephant

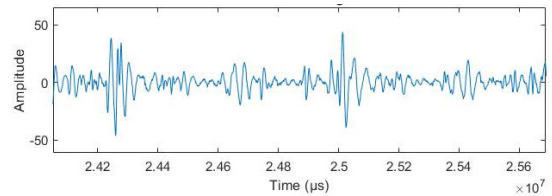


FIGURE 2. Seismic signal of elephants observed in our previous study [5].

footfall signal pattern from our previous study in [5], as illustrated in Fig. 2, which captures multiple consecutive footfalls generated by an elephant walking in a straight path with a typical gait.

2) MODIFIED ENERGY RATIO (MER)

The Modified Energy Ratio, an improved version of the energy ratio, enhances event detection by addressing challenges posed by varying signal amplitudes and background noise. According to [38], the energy ratio, $er[i]$, is defined at testing index i , where X represents the signal function and L denotes the window length in (2). Equation (3) defines the modified energy ratio at the test index i as $mer[i]$, extending the energy ratio by incorporating the instantaneous amplitude of the signal.

$$er[i] = \frac{\sum_{n_1=i}^{i+L} (X[n_1])^2}{\sum_{n_2=i-L}^i (X[n_2])^2} \quad (2)$$

$$mer[i] = [er[i] \cdot \text{abs}(X[i])]^3 \quad (3)$$

3) CONTEXTUALLY CUSTOMIZED WINDOW (CCW)

Compared to conventional event detection methods used in seismology, such as STA/LTA and MER, a detection method tailored to the specific characteristics of the target signal is expected to enhance both precision and efficiency. Accordingly, the CCW method was introduced as an experimental method built upon the shape of seismic signals generated by elephant footfalls.

As observed in our previous work [5] and illustrated in Fig. 2, a seismic signal produced by a single elephant footstep typically exhibits a higher amplitude concentrated at the center of the waveform. This feature is further corroborated by patterns reported in prior literature [6]. Based on this observation, in this study, a customized sliding window

²These samples were selected from Dt_Pinawala_Loc2.

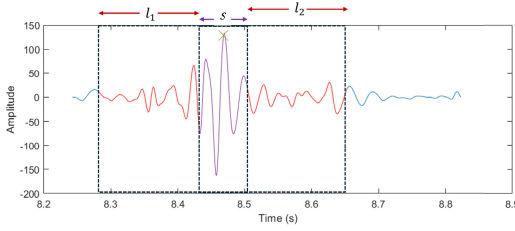


FIGURE 3. Window structure of CCW method.

structure was designed, consisting of a central short window flanked by two longer windows, as shown in Fig. 3.

In the single elephant footstep signal depicted in Fig. 3, the two flanking long windows, marked in red, have a length denoted by l_1 and l_2 , while the central short window, marked in purple, has a length denoted by s . Each window traverses the signal, and the ratio of average energy, $er_{ccw}[n]$, at index n is computed as introduced herein in (4). The signal function is represented by x .

$$er_{ccw}[n] = \frac{\frac{1}{s} \sum_{n_1=n-\frac{s}{2}}^{n+\frac{s}{2}} (x[n_1])^2}{\frac{1}{l_1} \sum_{n_2=n-l_1-\frac{s}{2}}^{n-\frac{s}{2}} (x[n_2])^2 + \frac{1}{l_2} \sum_{n_3=n+\frac{s}{2}}^{n+\frac{s}{2}+l_2} (x[n_3])^2} \quad (4)$$

As the sliding window aligns with the signal pattern of the elephant footstep as depicted in Fig. 3, a high, $er_{ccw}[n]$ is expected. In contrast, $er_{ccw}[n]$ remains relatively low for other signal patterns or continuous noise. Thus, potential seismic events corresponding to elephants can be effectively identified. Given the approximately equal lengths of l_1 and l_2 observed in Fig. 3, the subsequent application of the CCW method in this study employs the simplification $l_1 = l_2 = l$.

In the execution of the CCW method, the computation of the CCW energy ratio er_{ccw} , defined in (4), is performed as detailed in Algorithm 1. In this algorithm, a segment of the seismic signal with a sample length N is used; this length can be adjusted according to resource constraints but should exceed the duration of a single elephant footstep. The lengths of the analysis windows (s and l) are defined in terms of the number of discrete sampling points. The algorithm outputs the CCW energy ratio array (er_{ccw}). A similar moving window technique is employed for the STA/LTA and MER methods, with the respective ratio computations adapted according to (1) and (3).

To select the most suitable event detection method, the three aforementioned methods were compared in terms of accuracy and execution time. The results are presented in Table 6.

B. SELECTING AND EXTRACTING EVENTS

The selection of potential seismic events, while minimizing background noise, is achieved by applying three empirically defined thresholds to the outputs $er[n]$, $mer[n]$ and $er_{ccw}[n]$ corresponding to event detection algorithms STA/LTA, MER and CCW, respectively. The threshold for each method was

Algorithm 1 Optimized Computation of CCW Energy Ratio for Embedded Systems

Require: $X[1 \dots N]$: input seismic signal of length N
Require: s : short central window length
Require: l : flanking window length (for both sides)
Ensure: $er_{ccw}[]$: CCW energy ratio array

```

1:  $er_{ccw}[1 \dots N] \leftarrow$  array of zeros
2:  $s_{half} \leftarrow \text{floor}(s/2)$ 
3:  $n_{start} \leftarrow l + s_{half} + 1$ 
4:  $n_{end} \leftarrow N - s_{half} - l$ 
5:  $E_s, E_{l1}, E_{l2}$ : Variables to store average energy values
6: for  $n = n_{start}$  to  $n_{end}$  do
7:    $E_s \leftarrow 0$ 
8:    $E_{l1} \leftarrow 0$ 
9:    $E_{l2} \leftarrow 0$ 
10:  for  $i = n - s_{half}$  to  $n + s_{half}$  do
11:     $E_s \leftarrow E_s + X[i]^2$ 
12:  end for
13:   $E_s \leftarrow E_s / s$ 
14:  for  $i = n - s_{half} - l$  to  $n - s_{half} - 1$  do
15:     $E_{l1} \leftarrow E_{l1} + X[i]^2$ 
16:  end for
17:   $E_{l1} \leftarrow E_{l1} / l$ 
18:  for  $i = n + s_{half} + 1$  to  $n + s_{half} + l$  do
19:     $E_{l2} \leftarrow E_{l2} + X[i]^2$ 
20:  end for
21:   $E_{l2} \leftarrow E_{l2} / l$ 
22:   $er_{ccw}[n] \leftarrow E_s / (E_{l1} + E_{l2})$ 
23: end for
24: return  $er_{ccw}[]$ 

```

fine-tuned using reference video footage from field studies to ensure that the algorithm's output exceeded the threshold when a footfall event occurred.

Followed by, event detection, temporal segments with lengths outside the range defined by L_{Elp_Min} and L_{Elp_Max} were discarded to remove anomalies. Subsequently, event extraction was performed in three steps: (i) The midpoint index of each retained temporal segment was then designated as the point of interest (PoI), (ii) A constant bias value (L_{Bias}) was introduced to adjust the midpoint, with L_{Bias} which was optimized using video references. (iii) The event boundaries were computed using (5) and (6), corresponding to the start and end indices, respectively, with the extracted event defined by the segment between these indices.

$$\text{Starting index} = PoI + L_{Bias} - \frac{L_{Elp}}{2} \quad (5)$$

$$\text{Ending index} = PoI + L_{Bias} + \frac{L_{Elp}}{2} \quad (6)$$

The execution of the aforementioned process for event selection and extraction using the CCW method is detailed in Algorithm 2. The CCW energy ratio array (er_{ccw}) used in this algorithm is obtained from the preceding Algorithm 1. The same procedure can be applied to the STA/LTA and MER

Algorithm 2 Event Detection and Feature Extraction (CCW Method)

Require: $X[1 \dots N]$: input seismic signal
Require: $er_ccw[1 \dots N]$: CCW energy ratio array
Require: TH_{ccw} : threshold value for detection
Require: L_{Elp_Min} , L_{Elp_Max} : minimum and maximum allowable event lengths (in samples)
Require: L_{Bias} : bias offset for midpoint adjustment
Require: L_{Elp} : final fixed-length event segment

```

1:  $start\_flag \leftarrow \text{false}$ 
2: for  $n = 1$  to  $N - 1$  do
3:   if  $er\_ccw[n] < TH_{ccw}$  and  $er\_ccw[n + 1] \geq TH_{ccw}$  then
4:      $start\_index \leftarrow n$ 
5:      $start\_flag \leftarrow \text{true}$ 
6:   end if
7:   if  $er\_ccw[n] \geq TH_{ccw}$  and  $er\_ccw[n + 1] < TH_{ccw}$  and  $start\_flag$  then
8:      $end\_index \leftarrow n$ 
9:      $start\_flag \leftarrow \text{false}$ 
10:     $segment\_length \leftarrow end\_index - start\_index + 1$ 
11:    if  $L_{Elp\_Min} \leq segment\_length \leq L_{Elp\_Max}$  then
12:       $Pol \leftarrow \text{floor}((start\_index + end\_index)/2)$ 
13:       $Pol_{adj} \leftarrow Pol + L_{Bias}$ 
14:       $ExtractStart \leftarrow Pol_{adj} - \text{floor}(L_{Elp}/2)$ 
15:       $ExtractEnd \leftarrow Pol_{adj} + \text{floor}(L_{Elp}/2)$ 
16:       $Event \leftarrow X[ExtractStart : ExtractEnd]$ 
17:      // Perform feature extraction on the event
18:       $FeatureExtraction(Event)$ 
19:    end if
20:  end if
21: end for

```

methods by substituting the corresponding energy ratio arrays in place of er_{ccw} .

C. FEATURE EXTRACTION

Feature extraction transforms extracted seismic event into structured representations that capture the distinctive characteristics, enabling the differentiation of seismic events. In this study, nine features were extracted from the identified events.

These features are listed in Table 3, and are categorized into four feature types, representing signal characteristics in different domains: temporal, spectral, pattern matching, and statistical. Moreover, a reference term was assigned to each feature, which will be referred to hereinafter in this study.

1) TEMPORAL FEATURES

A fundamental temporal characteristic of a signal is the length of the event. In this study, event length represents the original length (in samples) of the detected event, as identified by the event detection method, prior to event extraction.

Moreover, motivated by its successful application in related studies [46], the number of zero crossings is also employed as a temporal feature. The number of zero crossings

TABLE 3. Features used in the study.

Feature Type	Feature	Reference Term
Temporal	Event Length (in number of samples)	Event Length
	Number of Zero Crossings	Zero Crossings
Spectral	Predominant Frequency	Pred. Frequency
	Maximum Cross-Correlation Value	Max Cross-Corr
Pattern Matching	Cross-Correlation Value at 0 Lag	Cross-Corr (0)
	Mean Squared Error	MSE
	DTW Alignment Cost	DTW
Statistical	Skewness	Skewness
	Kurtosis	Kurtosis

(Z) is computed using (7). Here, N denotes the length of the signal (in samples), $X[n]$ is the signal at sample n , and $\mathbb{I}$ is the indicator function. The indicator function returns 1 if $(X[n] \cdot X[n + 1]) < 0$ is true, and 0 otherwise [58].

$$Z = \sum_{n=1}^{N-1} \mathbb{I}[(X[n] \cdot X[n + 1]) < 0] \quad (7)$$

2) SPECTRAL FEATURES

A key characteristic of elephant footfalls is the low frequency, typically around 20 Hz [54]. Consequently, frequency-based features, such as the energy and correlation coefficients of power spectra, have been utilized to distinguish elephant footfalls from other signals [6], [46]. Notably, the predominant frequency was successfully employed to detect elephants in our previous study [5]. The predominant frequency is calculated using (8), where f_p represents the predominant frequency and $\mathcal{F}(X[n])$ is the Fourier Transform of the discrete signal $X[n]$, where n denotes the sample index in the time domain [59].

$$f_p = \arg \max_f |\mathcal{F}(X[n])|^2 \quad (8)$$

3) PATTERN MATCHING FEATURES

The features categorized under 'pattern matching' rely on a pre-identified and verified footfall signal pattern. This signal pattern was cross-referenced with video footage to ensure its accuracy. To address the variability in amplitude caused by differences in gain settings and detection range, the amplitude of each pattern was normalized before further analysis.

Cross-correlation is a method that has been successfully utilized for signal identification in seismology [60]. Therefore, it was also considered for identifying seismic events of elephant footfall. The cross-correlation at lag τ between the original signal $X[n]$ and the reference signal $Y[n]$ is represented in (9) as $R_{xy}(\tau)$, where N represents the length (in samples) of the original signal ($X[n]$) [61].

$$R_{xy}(\tau) = \sum_{n=0}^{N-1} X[n] \cdot Y[n + \tau] \quad (9)$$

However, for simplicity of implementation, the maximum cross-correlation value ($R_{\max}$) and the cross-correlation value at lag 0 ($R_{xy}(0)$) were used as features. These are represented

in (10) and (11), respectively.

$$R_{\max} = \max_{\tau} R_{xy}(\tau) \quad (10)$$

$$R_{xy}(0) = \sum_{n=0}^{N-1} X[n] \cdot Y[n] \quad (11)$$

Additionally, the extracted event was compared to the accurate signal pattern by calculating the mean squared error, which was also used as a feature. The mean squared error (*MSE*) is represented in (12), where $X[n]$ is the signal, $Y[n]$ is the reference signal pattern, and N is the length of both the original signal and the reference signal [62].

$$MSE = \frac{1}{N} \sum_{n=1}^N (X[n] - Y[n])^2 \quad (12)$$

Dynamic Time Warping (DTW) is a method introduced for time-series pattern recognition [63], and has also been applied to seismology-related applications [64]. DTW calculates the sum of distances between the optimally aligned points of two signal patterns. Thus, DTW alignment cost measures the difference between the extracted event and the reference footfall signal pattern (See Fig. 2), making it a potential feature. The DTW alignment cost ($DTW(X, Y)$) is computed over the optimal warping path P as shown in (13), where $X[i]$ is the original signal value at index i and $Y[j]$ is the reference signal value at index j [65].

$$DTW(X, Y) = \min_P \sum_{(i,j) \in P} \|X[i] - Y[j]\| \quad (13)$$

4) STATISTICAL FEATURES

Finally, the statistical features of skewness and kurtosis were used for classification, as inspired by [46]. Skewness measures the asymmetry of the distribution of the signal, while kurtosis quantifies the tailedness of the distribution. Equations (14) and (15) define skewness (S) and kurtosis (K), respectively. Here, $X[n]$ represents the recorded signal, N is the number of samples, μ is the mean of the signal, and σ is the standard deviation of the signal [66].

$$S = \frac{\frac{1}{N} \sum_{n=1}^N (X[n] - \mu)^3}{\sigma^3} \quad (14)$$

$$K = \frac{\frac{1}{N} \sum_{n=1}^N (X[n] - \mu)^4}{\sigma^4} \quad (15)$$

D. EVENT CLASSIFICATION

Two models namely, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) were chosen for the classification process. SVM was selected for its transparency and lower computational complexity, while ANN was employed as a reference model due to its potential for higher performance, despite its increased complexity.

The SVM classifier aims to optimize a hyperplane that maximizes the margin between the two classes while minimizing classification errors. Given a dataset consisting of N training samples, each sample is represented by a feature

vector $x_i \in \mathbb{R}^d$ and a corresponding class label $y_i \in \{-1, 1\}$. The optimization problem is defined as shown in (16), where W denotes the weight vector, b represents the bias term, C is the regularization parameter, and ξ_i are slack variables introduced to account for misclassifications [67], [68]. The constraints of the optimization problem are given in (17) [67].

$$\min_{W, b, \xi} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^N \xi_i \quad (16)$$

$$y_i(W^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, N. \quad (17)$$

For the SVM, several kernel functions including linear, third-degree polynomial, sixth-degree polynomial, Radial Basis Function (RBF), and sigmoid were evaluated. The kernel function maps the original input vectors into a higher-dimensional feature space, where linear separation becomes feasible. It is formulated in (18), where $K(x_i, x_j)$ denotes the kernel function, and $\phi(\cdot)$ represents a transformation that maps input vectors x_i and x_j into a higher-dimensional feature space [68].

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (18)$$

To enhance the model's generalization capability and to prevent overfitting, k -fold cross-validation was utilized. The cross-validation accuracy estimate is computed as the average of the individual accuracies obtained across all k folds [69]. Accordingly, the cross-validation accuracy (CV_k) given in (19), where k represents the number of folds, and $Accuracy_i$ denotes the accuracy obtained in the i^{th} fold. This method reduces variance and increases the reliability of the performance evaluation by averaging the results across multiple folds. In the conducted experiments, k was empirically set to 10.

$$CV_k = \frac{1}{k} \sum_{i=1}^k Accuracy_i \quad (19)$$

The ANN used in this study utilizes a sequential architecture tailored for binary classification, consisting of an input layer with 128 neurons, followed by three hidden layers with 64, 32, and 16 neurons, respectively. The Rectified Linear Unit (ReLU) activation functions were employed to enhance learning efficiency. The ReLU function was defined as (20), where x represents the input to the activation function [70].

$$\text{ReLU}(x) = \max(0, x) \quad (20)$$

The output layer contained a single neuron with a sigmoid activation function, mapping the network's output to a probability range of $[0, 1]$, making it suitable for binary classification. The sigmoid function ($\sigma(x)$) is given by (21), where x is the weighted sum of neuron input [71]. This function ensures that output can be interpreted as probabilities, facilitating the use of cross-entropy loss for training.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (21)$$

During feed-forward propagation, the weighted sum of inputs and biases at each layer was computed using (22) [72]. Here $z^{(l)}$ represents the pre-activation value at layer l , $W^{(l)}$ is the weight matrix, $a^{(l-1)}$ represents the activation values from the previous layer, and $b^{(l)}$ is the bias vector. The activation function was then applied to obtain the final activation value as expressed in (23), where $f(\cdot)$ represents the activation function [72], that is ReLU for hidden layers and sigmoid for the output layer.

$$z^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \quad (22)$$

$$\mathbf{a}^{(l)} = f(z^{(l)}) \quad (23)$$

To improve training stability and generalization, batch normalization was applied to the first two hidden layers, while dropout layers were incorporated to prevent overfitting. Specifically, a dropout rate of 30% was applied to the input and first hidden layer, while a 20% dropout rate was used in the second hidden layer.

The ANN model was trained using binary cross-entropy loss (L), which is a standard loss function for binary classification problems. It is defined in (24), where y_i is the true class label, and $\hat{p}_i$ is the predicted probability of the positive class for i^{th} sample [73].

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)] \quad (24)$$

The training was performed using backpropagation, where gradients were propagated backward through the network to update weights. The error term in each layer was computed as (25), where $\delta^{(l)}$ represents the error in layer l , $\frac{\partial L}{\partial a^{(l)}}$ is the derivative of the loss function (L) with respect to the activation, and $f'(z^{(l)})$ is the derivative of the activation function [74]. The gradient descent update rule of weight adjustment is given in (26), where η is the learning rate, $W^{(l)}$ represents the weight matrix in layer l , and $\frac{\partial L}{\partial W^{(l)}}$ is the gradient of the loss function (L) with respect to the weights [74]. The bias update was given by (27), where $b^{(l)}$ represents the biases in layer l , which were updated alongside the weights to optimize the network's performance [74].

$$\delta^{(l)} = \frac{\partial L}{\partial \mathbf{a}^{(l)}} \odot f'(z^{(l)}) \quad (25)$$

$$W^{(l)} = W^{(l)} - \eta \frac{\partial L}{\partial W^{(l)}} \quad (26)$$

$$b^{(l)} = b^{(l)} - \eta \delta^{(l)} \quad (27)$$

For further optimization, the Adam optimizer was employed with a reduced learning rate of 0.0005 to ensure stable convergence. Adam optimizer adaptively adjusts the learning rate for each parameter using first and second-moment estimates. The weight update rule for Adam optimizer is given in (28), where w_t is the weight at iteration t , η is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second-moment estimates of the gradients, and ϵ is a small

constant [75].

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (28)$$

For training the neural network, 80% of each dataset was allocated while 20% was reserved for testing. The selected models were trained under two distinct scenarios: (i) using a dataset from a relatively controlled environment, (ii) from a more natural setting influenced by environmental factors such as wind. Following training, the models were tested under three different conditions: (i) for a controlled environment, (ii) for wild elephants in their natural habitat, and (iii) for a more complex scenario involving wild elephants in both natural and human-inhabited areas, closely resembling HEC-prone environments.

To evaluate the performance of all of these models under different scenarios, both accuracy and the F1 score were calculated. Accuracy measures the proportion of correctly classified instances relative to the total number of instances as in (29). Here, TP , TN , FN , and FP are confusion matrix components, which are defined in Table 4

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (29)$$

To provide a more comprehensive evaluation of the model's performance, the F1-score was utilized. The F1-score represents the harmonic mean of precision and recall, offering a balanced measure between false positives and false negatives. Precision is defined in (30) and recall is computed in (31). The F1-score is then calculated as shown in (32), ensuring a robust evaluation of the classification model.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (30)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (31)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (32)$$

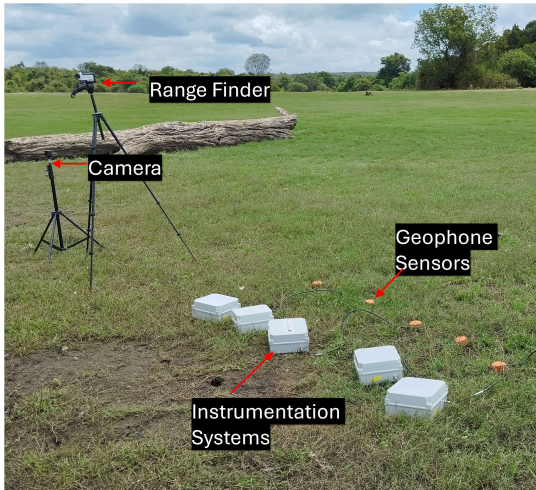
Finally, the ANN model was used for explainable AI analysis to provide interpretability to the classification results. To this end, SHapley Additive exPlanations (SHAP) values were calculated using (33) [76]. Here, ϕ_i represents the contribution of feature i to the model's prediction, S is a subset of features excluding i , and $f(s)$ is the output when using only the features in S .

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup i) - f(S)] \quad (33)$$

These values allowed for identifying the contribution of each feature to the model's performance. This step was included to determine the most important features for elephant footfall identification. As such, in future implementations, this knowledge could be used to optimize the classification process by pruning less important features, thereby improving efficiency in resource-constrained implementations (i.e., embedded systems).

TABLE 4. Confusion matrix metrics and their interpretations.

Confusion Metric	Definition	Classification Scenario
TP	True Positive	Correctly classifies an elephant signal as an elephant
TN	True Negative	Correctly classifies a non-elephant signal as not from an elephant
FP	False Positive	Misclassifies a non-elephant signal as from an elephant
FN	False Negative	Misclassifies an elephant signal as not from an elephant

**FIGURE 4.** Experimental setup used in field studies.

E. SIGNAL ACQUISITION SETUP

Acquiring seismic signals from elephant footfalls was challenging, particularly over long distances, due to their very low amplitudes and high susceptibility to noise [44]. To address this challenge, this study employed the instrumentation system that was introduced in our previous paper [5], which incorporates a novel geophone-sensor interface.

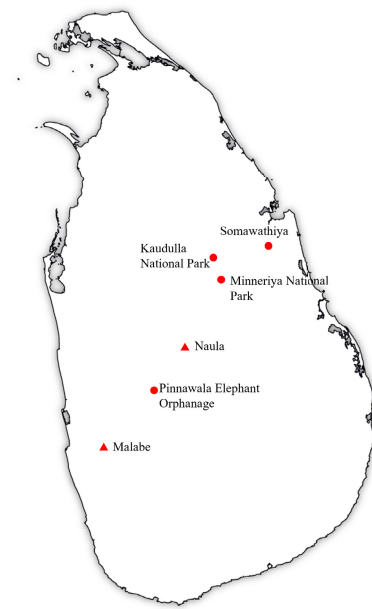
The geophone-sensor interface is specifically designed to enhance the detection of elephant footfall signals by amplifying the desired signals while suppressing ambient noise. The system comprises a high-sensitivity (85.8 V/m/s) 10 Hz vertical geophone, a microcontroller for signal formatting and buffering, and a battery backup to ensure continuous operation. Additionally, a Raspberry Pi module is utilized for real-time signal acquisition.

Based on our empirical knowledge from previous study [5], the systems were configured with three distinct gain settings: $S_0 = 35$, $S_1 = 580$, and $S_2 = 25,911$. Data acquisition was conducted at an approximate sampling frequency of 880 Hz.

The complete experimental setup used in the field studies is illustrated in Fig. 4. In this setup, based on the experimental scenario, multiple sensors, ranging from one to five, were deployed for seismic signal acquisition. Furthermore, each experiment was recorded using a camera setup to capture reference videos. Additionally, a range finder was employed to accurately measure the distances between the sensor placements and the elephants.

F. FIELD STUDIES FOR DATA COLLECTION

Using the signal acquisition setup, extensive field studies were conducted at multiple locations across Sri Lanka,

**FIGURE 5.** Data collection sites (• - Sites for the collection of elephant datasets, ▲ - Sites for the collection of non-elephant datasets).

as illustrated in Fig. 5. To enhance model generalization, as shown in this figure, elephant-induced seismic signals were acquired from multiple HEC-prone regions across the country. Additionally, to improve model accuracy in real-world deployments, the non-elephant seismic dataset incorporated ambient seismic activities commonly observed in HEC-affected areas.

Field studies involving elephants were carried out under the strict supervision of the Department of National Zoological Gardens and the Department of Wildlife Conservation, Sri Lanka. All studies involving human participants were conducted with informed consent following a full explanation of the procedures.

First, seismic signal datasets of elephant footfalls was collected under controlled conditions at the Pinnawala Elephant Orphanage, located in Rambukkana, Sri Lanka. The data acquisition involved three partially tamed elephants allowed to move freely across two designated paths: Location 1, a 100-meter stretch, and Location 2, a 160-meter path. The maximum recorded detection range was 155.6 meters at Location 2. Both sites were relatively isolated from anthropogenic activity and external seismic disturbances, enabling the collection of high-quality seismic data with minimal noise interference.

Second, seismic signal datasets of elephant footfalls in natural habitats was collected at two sites in Sri Lanka:



FIGURE 6. Sensor and instrumentation system placement in the field study at Kawudulla.

Kawudulla National Park ($8^{\circ}09'52.8''\text{N}$, $80^{\circ}54'47.4''\text{E}$) and Minneriya National Park ($8^{\circ}00'57.4''\text{N}$, $80^{\circ}51'20.8''\text{E}$). At both sites, a large group of wild elephants was observed strolling and feeding in a grassland area bordered by forest and a reservoir. In Kawudulla, seismic signals were recorded from wild elephants at distances ranging from 66.9 meters to 140 meters, while in Minneriya, the observed range was between 85.3 meters and 106 meters.

The positioning of the instrumentation system with the connected geophone sensors relative to the movement paths of the herd of elephants at the aforementioned long ranges is depicted in Fig. 6. Even at these challenging distances, where visual confirmation was limited due to dense vegetation, the setup enabled effective seismic signal acquisition.

Third, datasets of wild elephants intruding into a human-inhabited area were collected in Somawathiya, Sri Lanka, across two distinct locations within a settlement bordering a forest reserve. In Location 1, seismic signals were recorded from a single wild elephant near an electric fence, with the sensor placed in close proximity to the animal, as shown in Fig. 7. However, this dataset was affected by interference from human activities and vehicular movements. Location 2 was a relatively quiet riverbank within the same human-inhabited area, where sensors were positioned approximately 32 meters from the observed elephant. Notably, no aggressive behavior was exhibited by the elephants during either encounter.

Moreover, non-elephant datasets comprise human footfalls, recorded in the outdoor premises of the Sri Lanka Institute of Information Technology. The dataset includes three subjects, both male and female, walking along a 15-meter straight path. Seismic signals from a moving motorcycle were also recorded at the same location.

To resemble non-elephant activity typically encountered in HEC-prone village environments, signal datasets were collected in Melpitiya, Naula, Sri Lanka ($7^{\circ}41'33.8''\text{N}$, $80^{\circ}37'39.3''\text{E}$), within a paddy field adjacent to a village. In the first phase, signals were captured from the footfalls of roaming cattle. In the second phase, a combination of human

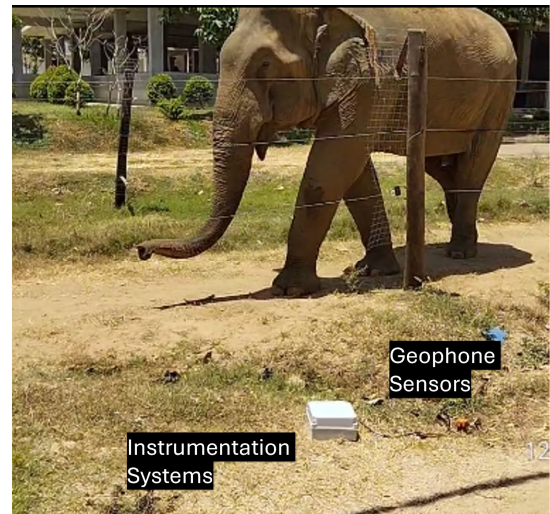


FIGURE 7. Sensor and instrumentation system placement in the field study at Somawathiya).

and cattle footfalls, along with occasional dog footfalls, was recorded. This resulted in an unsupervised dataset that closely reflects real-world non-elephant disturbances in HEC-affected areas.

Notably, the field studies conducted in Kawudulla, Minneriya, and Naula were subjected to considerable wind conditions, while Somawathiya experienced mild wind. The intermittent movement of grass and paddy due to this wind conditions may have introduced disturbances in the recorded seismic signals.

Based on the data collected from the aforementioned studies, the datasets presented in Table 5 were prepared. The table shows the corresponding field study, the reference name for each dataset, and the gain settings used for the selected recordings.

IV. RESULTS AND DISCUSSION

The analysis and pre-processing of seismic signals, along with the aforementioned methodologies for event detection and feature extraction, were conducted using MATLAB R2018a on a computer equipped with an Intel Core i5-1035G1 processor (1.00 GHz) and 8 GB of RAM. Subsequent classification and explainable AI (XAI) models were implemented in Python 3 on Google Colab within a CPU-only environment. The following subsections present the results corresponding to each methodological component.

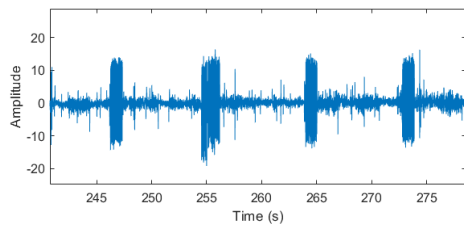
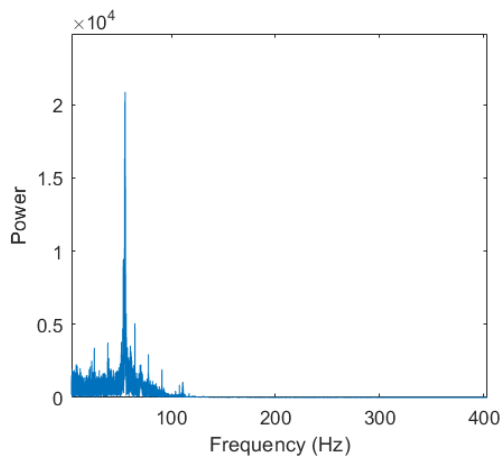
A. IMPACT OF AMBIENT NOISE

Across all field studies, the instrumentation system successfully recorded seismic signal patterns under the given gain settings. To attenuate high-frequency noise while minimizing distortion, a 6th-order low-pass Butterworth filter with a cutoff frequency of 80 Hz was applied initially.

However, as shown in Fig. 8 (corresponding power spectrum in Fig. 9), it can be observed that data gathered from

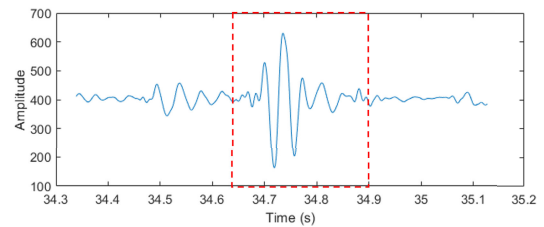
TABLE 5. Summary of experiment sites and corresponding datasets.

Experiment Site	Scenario	Subject	Dataset Ref.	Gain Settings
Pinnawala	Location 1	Elephant (Controlled)	Dt_Pinnawala_Loc1	S1
Pinnawala	Location 2	Elephant (Controlled)	Dt_Pinnawala_Loc2	S0, S1, S2
Kawudulla	-	Elephant (Wild)	Dt_Kawudulla	S3
Minneriya	-	Elephant (Wild)	Dt_Minneriya	S3
Somawathiya	Location 1	Elephant (Wild)	Dt_Somawathiya_Loc1	S3
Somawathiya	Location 2	Elephant (Wild)	Dt_Somawathiya_Loc2	S3
Malabe	-	Human	Dt_Human	S3
Malabe	-	Motorcycle	Dt_Vehicle	S3
Naula	Phase 1	Cattle	Dt_Cattle	S3
Naula	Phase 2	Combination	Dt_Comb	S3

**FIGURE 8.** Time-domain representation of noise observed in the field study at Kawudulla.**FIGURE 9.** Power spectrum representing the noise observed in the field study at Kawudulla.

Kawudulla, Minneriya, and Naula exhibit additional noise, characterized by an abnormal peak at approximately 55 Hz. A similar noise artefact was also detected in the Somawathiya field study, albeit at a lower amplitude. With the support of the video references, it was identified that the noise was caused by the periodic movement of vegetation induced by wind.

To mitigate the identified noise, a band-stop filter was applied with stopband frequencies at 50 Hz and 60 Hz. The filter's order and cutoff frequencies were designed using Butterworth parameters to achieve a 60 dB attenuation, effectively suppressing the targeted noise. The recovered signal from Kawudulla, Minneriya, and Somawathiya then provided the expected pattern for elephant footfalls, as shown in Fig. 10.

**FIGURE 10.** Seismic signal pattern of elephant footfall.**FIGURE 11.** Footfall of elephants.

B. FOOTFALL PATTERN OF ELEPHANTS

In relatively controlled settings, where ambient noise is less, the repetitive footfall signal pattern of an elephant is depicted in Fig. 10. A closer inspection of this pattern reveals that the signal contains two distinct patterns with varying amplitudes, likely resulting from the movement of both the front and rear legs (see Fig. 11). This hypothesis aligns with findings in [6], which confirms that the footfall from the front leg generates a larger amplitude compared to the rear leg. However, in a noisy environment, only the segment corresponding to the front leg could be frequently observed (the segment highlighted in Fig. 10). Therefore, this segment was considered as the characteristic footfall signature of elephants for the remainder of the study.

After manually inspecting the events, the average length of the considered footfall signature of elephants was found to be approximately 190 data samples. As the sampling frequency is 880 Hz for the setup, the average temporal length of the signal L_{Elp} is calculated as 215.90 ms. This aligns with previous studies such as [54], which has reported that the typical elephant footfall length ranges from 103 ms to 250 ms. Based on the observations in this study, the minimum and maximum footfall durations denoted as L_{Elp_Min} and L_{Elp_Max} are considered to be 75.00 ms and 354.55 ms, respectively. These parameters were utilized in event detection.

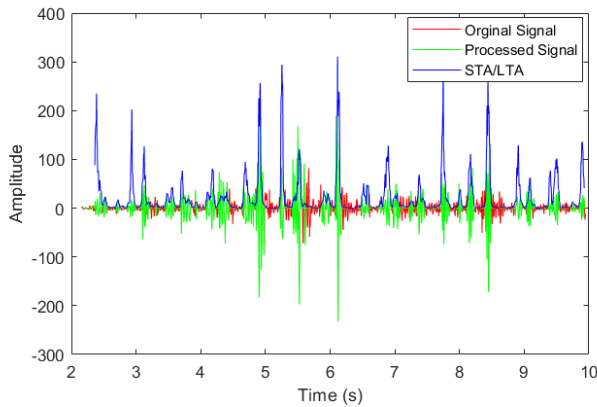


FIGURE 12. Results of STA/LTA method.

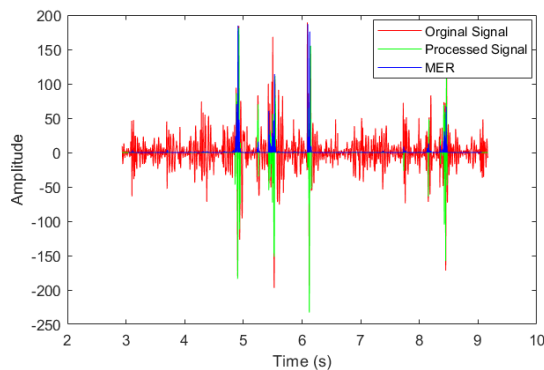


FIGURE 13. Results of MER method.

TABLE 6. Comparison of event detection methods based on detection accuracy and execution time.

Method	Detected	Missed	Merged	Execution Time (ms)
STA/LTA	10	0	2	13.84
MER	4	6	0	14.42
CCW	10	0	1	16.48

C. EVALUATION OF EVENT DETECTION METHODS

To evaluate the performance of event detection methods, a 5.2-second temporal segment of the signal recording was selected, and the corresponding video footage within this time frame was analyzed. Within this segment, ten footsteps were visually identified. These events were cross-referenced and confirmed with the identifiable footsteps in the video footage. The confirmed footsteps within this segment were used as the ground truth.

Subsequently, this segment was processed using three event detection methods: STA/LTA, MER, and CCW. The results from each event detection method were obtained, as illustrated in Fig. 12, Fig. 13, and Fig. 14, which present the corresponding outputs. In these figures, the time span from 3.4 s to 8.6 s represents the selected temporal segment. A comparison of the results is provided in Table 6.

The detection performance of each event detection method for the selected time span is represented in Table 6.

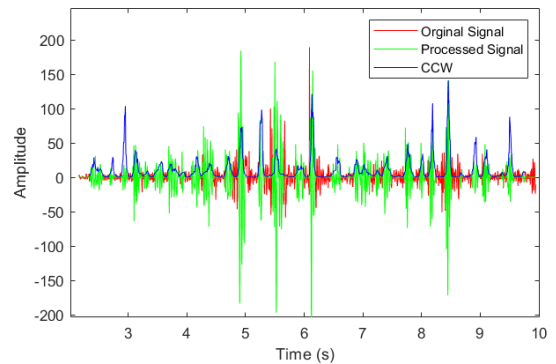


FIGURE 14. Results of CCW method.

Accordingly, both the STA/LTA and CCW methods successfully detected all the identified footsteps. However, the MER method was unsuccessful, missing six events. During the detection process, there were instances where the signal was captured but merged with additional seismic components. These events were recorded under the column titled ‘Merged’. A merged event can introduce deviations in the *Pol* during event extraction, leading to inaccurate results in subsequent processing. Within the limited time frame, STA/LTA detects two instances, while CCW identifies one, leading to the hypothesis that CCW is superior in event separation. However, it must be noted that even minor irregular foot movements by elephants can generate unpredictable seismic patterns. As a result, the authors are unable to confirm false detections in these scenarios, and false detections were not considered as factors for evaluating performance.

The average execution time of each method was included as a criterion for evaluating performance (see Table 6). Among the three methods, the STA/LTA method demonstrated the lowest execution time. Although the CCW method showed acceptable results, an assessment of both detection accuracy and execution time confirmed that the STA/LTA method exhibited the desirable performance in the given context of constrained computing resources. Consequently, it was selected for further processing.

Using the STA/LTA method, potential events were identified and extracted from all the datasets listed in Table 5. As outlined in the selecting and extracting events (refer subsection III-B), events with a temporal length shorter than, L_{Elp_Min} (75.00 ms) or exceeding L_{Elp_Max} (354.55 ms) were rejected.

D. RESULTS OF CLASSIFICATION

Following a comprehensive examination of events extracted from the field study in Pinnawala, specifically from the datasets Dt_Pinnawala_Loc1 and Dt_Pinnawala_Loc2 (refer Table 5), each elephant footstep was manually annotated with video references to ensure accuracy. However, in field studies conducted in Kawudulla, Minneriya and Somawathiya, precise identification of each individual

TABLE 7. Number of samples corresponding to each dataset.

Dataset Reference	Number of Samples
Dt_Pinnawala_Loc1	96
Dt_Pinnawala_Loc2	132
Dt_Kawudulla	41
Dt_Minneriya	14
Dt_Somawathiya_Loc1	23
Dt_Somawathiya_Loc2	9
Dt_Human	92
Dt_Vehicle	7
Dt_Cattle	750
Dt_Comb	767

TABLE 8. Summary of training and testing cases with corresponding datasets.

Case Name	Class	Dataset Used
Train_Case_1	Positive	Dt_Pinnawala_Loc2
	Negative	Dt_Human Dt_Cattle Dt_Vehicle
Train_Case_2	Positive	Dt_Pinnawala_Loc2
	Negative	Dt_Human Dt_Vehicle
Test_Case_1	Positive	Dt_Pinnawala_Loc1
	Negative	Dt_Comb
Test_Case_2	Positive	Dt_Kawudulla
	Negative	Dt_Comb
Test_Case_3	Positive	Dt_Kawudulla Dt_Minneriya Dt_Somawathiya_Loc1 Dt_Somawathiya_Loc2
	Negative	Dt_Comb

footstep was not achievable due to obstacles that hindered close observation in the natural environment. Instead, seismic signals from one-minute time intervals where footfalls were visible in the video footage were annotated. Furthermore, these annotations were validated through manual verification and cross-referenced with the pattern illustrated in Fig. 11. All extracted events from field studies conducted in Malabe and Naula (events from non-elephant seismic datasets) were subsequently annotated as ‘non-elephant events’.

The number of samples (extracted seismic events) of each dataset is presented in Table 7.³ It was observed that the Dt_Cattle dataset exhibited a substantially higher number of recorded events, posing a potential risk of class imbalance during model training. To address this, random under-sampling was applied to this dataset to align with the mean event count of the other datasets.

The values for the feature specified in Section 3 were computed for each extracted event. These features were then systematically categorized as specified in Table 8. Train_Case_1 and Train_Case_2 were used to train two distinct versions of each classifier. Train_Case_1 included a negative dataset, Dt_Cattle, containing noise disturbances such as wind and grass movement, which are common in HEC-prone natural environments in Sri Lanka.

³For these experiments, only precisely matched elephant footfalls were selected from identified elephant footfall events.

In contrast, Train_Case_2 comprised a dataset from a relatively noise-free setting. Test_Case_1, Test_Case_2, and Test_Case_3 were used to evaluate the trained classifiers, representing the following scenarios: (1) partially tamed elephants in controlled settings, (2) wild elephants in their natural habitats, and (3) wild elephants in both natural and human-influenced environments.

The classification was performed as outlined in the methodology section, using both ANN and SVM (refer III-D). The SVM was trained using five different kernels. Each model was trained on two distinct versions of the dataset. The classification results are presented in Table 9.

In Table 9, for SVM, the ‘Test Dataset’ column shows performance across all test folds, obtained using k-fold cross-validation, where each fold serves as a validation set in successive iterations. For ANN, it represents the model’s performance on the 20% subset of the dataset reserved for evaluation after training. Test_Case_1, Test_Case_2, and Test_Case_3 correspond to the evaluation of the classifiers under the respective scenarios described earlier.

According to Table 9, the performance evaluation using the Test Dataset demonstrated that the ANN model, serving as the reference benchmark, achieved the highest performance with an accuracy and F1 score of 100 %. This outcome aligns with expectations, given ANN’s capacity to model complex patterns and fit the training data effectively. Among the SVM models, the SVM with an RBF kernel exhibited the best performance, achieving 99 % accuracy and F1 score when trained on Train_Case_2 (Model B4), compared to 89 % accuracy and F1 score for Train_Case_1 (Model A4). This finding highlights that the controlled settings generally lead to superior performance due to reduced noise interference.

In Test_Case_1, which evaluates model performance in a controlled test setting, the ANN trained on Train_Case_2 (Model B6) outperformed the ANN trained on Train_Case_1 (Model A6), reinforcing the earlier observation regarding the benefits of controlled conditions. However, despite this advantage, ANN experienced a substantial drop in accuracy and F1 score compared to the Test Dataset, suggesting that its higher complexity may have led to overfitting, reducing its robustness in varying conditions.

Although SVM models also exhibited a performance degradation in Test_Case_1, the decrease was notably less severe compared to ANN. In terms of accuracy, the SVM with a Linear kernel (Models A1 and B1) marginally outperformed the RBF kernel. However, when considering the F1 score, both SVM with Linear and RBF kernels achieved 85 %, the highest among all models, confirming that the SVM with an RBF kernel maintained satisfactory performance.

When both models were applied to elephant data in natural environments in Test_Case2, the ANN model trained on Train_Case_1 (Model A6) outperformed the model trained with Train_Case_2 (Model B6), which

TABLE 9. Classification results for STA/LTA as an event detection method.

Model ID	Classifier Used	Training Case	Test Dataset		Test Case 1		Test Case 2		Test Case 3	
			Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
A1	SVM (Kernel: Linear)	Train_Case_1	89%	89%	81%	85%	69%	56%	67%	74%
A2	SVM (Kernel: Poly, degree=3)		86%	86%	73%	81%	64%	27%	70%	79%
A3	SVM (Kernel: Poly, degree=6)		77%	79%	72%	81%	62%	23%	64%	76%
A4	SVM (Kernel: RBF)		90%	90%	80%	85%	73%	60%	70%	76%
A5	SVM (Kernel: Sigmoid)		85%	85%	75%	80%	65%	54%	59%	66%
A6	ANN		100%	100%	70%	70%	66%	54%	64%	61%
B1	SVM (Kernel: Linear)	Train_Case_2	99%	100%	81%	85%	69%	56%	67%	74%
B2	SVM (Kernel: Poly, degree=3)		100%	100%	73%	81%	64%	27%	70%	79%
B3	SVM (Kernel: Poly, degree=6)		82%	86%	72%	81%	62%	23%	64%	76%
B4	SVM (Kernel: RBF)		99%	99%	80%	85%	73%	60%	70%	76%
B5	SVM (Kernel: Sigmoid)		100%	100%	75%	80%	65%	54%	59%	66%
B6	ANN		100%	100%	78%	83%	56%	55%	60%	66%

was the opposite for the relatively controlled settings. This outcome suggests that the inclusion of the *Dt_Cattle* dataset enhanced the model performance. It is hypothesized that natural phenomena, such as wind or resulting grass movement in corresponding field studies contributed to improving the model's generalization capability.

In contrast, the performance of the SVM with the RBF kernel remained consistent regardless of the training dataset; indicating that SVM primarily relies on support vectors to define decision boundaries, making it less sensitive to dataset variations and more reliant on fundamental feature separability. Notably, the SVM model with the RBF kernel yielded the best results in the natural environment with an accuracy of 73% and F1 score of 60%.

In comparison with [50] (mentioned in Table 1), a similar experiment conducted in natural environments for Asian elephants using dedicated embedded systems, this study was able to detect elephants from over 140 m, whereas [50] was limited to short detection ranges such as 50 m.⁴ Thereby surpassing existing range benchmarks for dedicated embedded systems.

Moreover, when comparing this study to a study that employed high-end seismometers [51] and reported a comparable accuracy of 77% at a range of 150 m, this work, even using relatively low-cost, custom embedded systems, achieved a comparable result with only a 4% reduction in accuracy at a similar range of 140 meters. These results therefore demonstrate a competitive balance between accuracy, detection range, and cost-effectiveness, supporting the authors' motivation to implement this approach as a viable solution for the HEC problem in natural environments. The results indicate competitive performance relative to prior studies, summarized in Table 1.

In *Test_Case_3*, the most challenging test case, ANN which was trained on *Train_Case_1* (Model A6) (accounted for natural phenomena), achieved 64% accuracy

⁴The exact detection range was not specified; however, sensors were placed 50 meters from electrified railway lines, the expected crossing points of elephants.

and an F1 score of 76%. However, despite the difficult conditions, the SVM with the RBF kernel (Models A4 and B4) outperformed ANN, achieving 70% accuracy and an F1 score of 76%, the highest among all models for this test case. This result underscores the robustness of the SVM model in complex environments. Most importantly, this is the first study to experiment with seismic signal classification on wild Sri Lankan elephants that entered human-inhabited areas, closely resembling real-world HEC scenarios.

For further analysis, the execution times of the SVM with the RBF kernel and ANN were compared across three test cases and both training sets. The ANN model exhibited an average execution time of 230.7 ms, whereas the SVM with the RBF kernel required only 147.69 ms. Overall, the SVM with the RBF kernel achieved high classification performance while maintaining a significantly lower execution time, reinforcing its suitability for deployment in resource-constrained implementations.

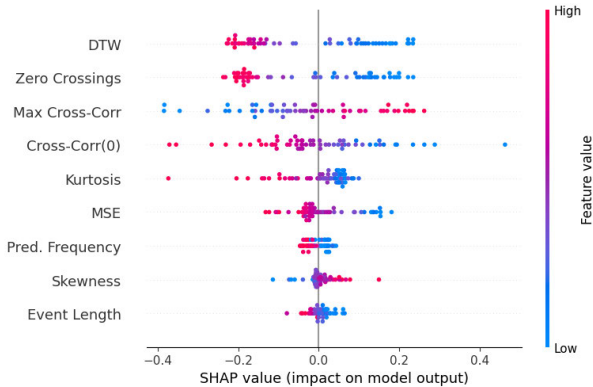
In summary, the ANN model performed well under its trained conditions but struggled to generalize, especially for natural environments. The SVM with the RBF kernel demonstrated stable performance across all test scenarios, including those closely resembling HEC-prone areas. Consequently, while ANN excels under ideal training conditions, SVM with the RBF kernel proves to be more reliable for practical deployment in a real-world environment.

E. PERFORMANCE AND FEASIBILITY OF CCW

CCW demonstrated comparatively accurate event detection (refer Table 6), prompting further evaluation of its classification performance as an event detection method. The classification outcomes using CCW were comparable to those obtained with the STA/LTA method, as shown in Table 10. The comparison includes the best-performing classifier: SVM with RBF kernel, and the reference model: ANN. Each model was trained and tested using *Train_Case_1* and *Train_Case_2* with both CCW and STA/LTA as event detection methods.

TABLE 10. Classification performance using CCW and STA/LTA event detection methods.

Classifier	Training Case	CCW		STA/LTA	
		Accuracy	F1 Score	Accuracy	F1 Score
SVM (RBF Kernel)	Train Case 1	88%	87%	90%	90%
	Train Case 2	99%	99%	99%	99%
ANN	Train Case 1	92%	92%	100%	100%
	Train Case 2	100%	100%	100%	100%

**FIGURE 15.** Summary plot for Train_Case_1.

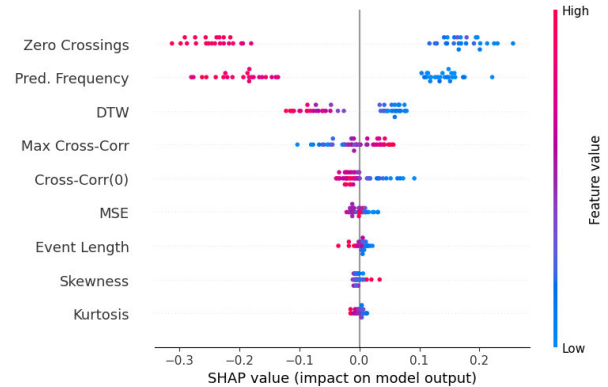
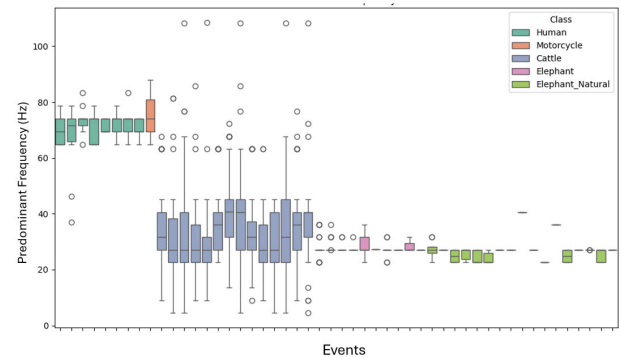
The comparative analysis in Table 10 indicates that both event detection methods yield nearly identical classification performance, particularly under Train_Case_2. The CCW method demonstrates exceptionally high accuracy and F1 scores, achieving 99% for SVM with an RBF kernel and 100% for ANN, aligning with the results obtained using STA/LTA. These findings confirm that CCW is as effective as STA/LTA in relatively controlled settings.

However, under the more challenging Train_Case_1, a slight decline in performance is observed with the CCW method relative to STA/LTA. While the ANN model exhibits an 8% reduction in both accuracy and F1 score, the performance drop in the more generalized SVM model is relatively minor, with accuracy decreasing by only 2% and F1 score by 3%. This variation is expected in complex natural scenarios leading to a conclusion that environmental factors may introduce additional uncertainties.

The slight reduction in performance observed under Train_Case_1 is likely to be caused by the fact that the Dt_Cattle dataset in Train_Case_1 contains non-elephant events with structural similarities to the pattern⁵ in Fig. 3. As the CCW method is still in its early stages, further investigations are necessary to confirm these assumptions and to enhance its robustness across diverse datasets which will be a significant future work of this study.

However, it is worth mentioning that even in its early stages, the CCW method demonstrated competitive classification performance compared to other event detection approaches. Despite its increased computational complexity,

⁵A similar anomaly is further investigated under feature impact assessment.

**FIGURE 16.** Summary plot for Train_Case_2.**FIGURE 17.** Distribution of predominant frequency for each event.

preliminary experimental results indicate that CCW outperformed existing methods in specific aspects, particularly in its resilience to event merging (refer to Table 10). These findings highlight the strong potential of CCW for real-world applications. Further optimization of the algorithm and fine-tuning of its parameters could enhance classification accuracy and improve its feasibility for deployment in embedded systems.

F. FEATURE IMPACT ASSESSMENT

Considering the high accuracy achieved during validation on the training dataset and the observed performance variations in natural environments, XAI with SHAP was utilized to analyze the contribution of each feature to the referenced ANN model. The summary plot for Train_Case_1 is shown in Fig. 15, while that for Train_Case_2 is presented in Fig. 16.

As shown in Fig. 15 and Fig. 16, the number of zero crossings, DTW alignment cost exhibit a significant impact on classification. Moreover, Predominant Frequency; a widely used parameter with high significance in related studies [5], [6] demonstrated strong influence under relatively controlled settings (Fig. 16) but showed a notably reduced impact when trained under more natural conditions (Fig. 15).

This reduction in the impact of Predominant Frequency may be attributed to the influence of non-elephant events that fall within a similar frequency range to elephant footfalls due to local soil conditions and environmental phenomena. This effect was further analyzed through the distribution of Predominant Frequency for both elephant and non-elephant events, as illustrated in Fig. 17. It shows that the Predominant Frequency of cattle footfalls sometimes aligns with the predominant frequency range of elephant footfalls, which is suspected to occur due to wet soil conditions. However, further research is necessary to confirm whether these conditions are specific to Sri Lankan HEC-prone locations, such as paddy fields.

Despite these challenging conditions, features such as the number of zero crossings and DTW alignment cost maintain high classification performance across different environments. Interestingly, features like event length and skewness have a negligible influence on classification. Thus, removing these features can lower computational costs without significantly affecting the overall accuracy.

V. CONCLUSION

This study presented a framework to automate the manual classification process in elephant footfall detection and classification. Between the event detection methods compared in this study, STA/LTA method was identified as the most suitable for resource-constrained implementations, incurring the CCW: the novel method presented in this study, tailored specifically for elephant footfall detection requires future attention. However, even at its early stage, the CCW method demonstrated competitive performance with the well-established STA/LTA approach and effectively mitigated the event-merging effect.

The results show that elephant footfalls from 156.6 m away were detected and classified under controlled conditions, whereas, footfalls were detected up to 140 m away in natural environments. The model achieved a footfall detection accuracy of 99% in controlled settings and 73% in natural environments, which is among the highest reported in comparable studies. Yet there is a 3% accuracy drop in the most challenging environments: human habitats prone to HEC. However, the overall performance of the framework in terms of accuracy, range, cost-effectiveness, and practical implementation, concludes that it is a promising solution for long-range elephant detection. Additionally, it was identified that there is an impact on relatively simple features such as DTW alignment cost and number of zero-crossings for classification. Importantly, DTW alignment cost has not

been previously explored in related research, underscoring its novelty and potential value.

In future, the CCW method will also be extended to operate beyond elephant footfall detection, offering potential benefits for various ambient intelligence applications. Furthermore, the instrumentation system will be enhanced with automatic gain adjustment to improve accuracy across a broader range of detections.

ACKNOWLEDGMENT

The authors express their profound gratitude to Mendis Wickremasinghe for his invaluable guidance throughout this research. Sincere appreciation is also extended to Major General Lal Gunasekara for his strategic planning and facilitation of field studies. Furthermore, they acknowledge the Department of Wildlife Conservation, Sri Lanka, and the Department of National Zoological Gardens, for their support, guidance, and granting of necessary permissions to conduct field studies within their premises.

REFERENCES

- [1] L. J. Shaffer, K. K. Khadka, J. Van Den Hoek, and K. J. Naithani, "Human-elephant conflict: A review of current management strategies and future directions," *Frontiers Ecol. Evol.*, vol. 6, pp. 1–12, Jan. 2019.
- [2] Z. M. Thant, R. May, and E. Røskaft, "Pattern and distribution of human-elephant conflicts in three conflict-prone landscapes in Myanmar," *Global Ecol. Conservation*, vol. 25, Jan. 2021, Art. no. e01411.
- [3] C. Santiapillai, S. Wijeyamohan, G. Bandara, R. Athurupana, N. Dissanayake, and B. Read, "An assessment of the human-elephant conflict in Sri Lanka," *Ceylon J. Sci., Biol. Sci.*, vol. 39, no. 1, pp. 21–33, Oct. 2010.
- [4] T. G. S. L. Prakash, A. W. Wijeratne, and P. Fernando, "Human-elephant conflict in Sri Lanka: Patterns and extent," *Gajah*, vol. 51, pp. 16–25, Jan. 2020.
- [5] J. L. Wijayaraja, J. L. Wijekoon, M. Wijesundara, and L. J. M. Wickramasinghe, "Toward long range detection of elephants using seismic signals: A geophone-sensor interface for embedded systems," *IEEE Access*, vol. 12, pp. 72210–72223, 2024.
- [6] J. D. Wood, C. E. O'Connell-Rodwell, and S. L. Klemperer, "METHOD-OLOGICAL INSIGHTS: Using seismic sensors to detect elephants and other large mammals: A potential census technique," *J. Appl. Ecol.*, vol. 42, no. 3, pp. 587–594, Jun. 2005.
- [7] B. T. Arnason, L. A. Hart, and C. E. O'Connell-Rodwell, "The properties of geophysical fields and their effects on elephants and other animals," *J. Comparative Psychol.*, vol. 116, no. 2, pp. 123–132, 2002.
- [8] A. Szenicer, M. Reinwald, B. Moseley, T. Nissen-Meyer, Z. Mutinda Muteti, S. Oduor, A. McDermott-Roberts, A. G. Baydin, and B. Mortimer, "Seismic savanna: Machine learning for classifying wildlife and behaviours using ground-based vibration field recordings," *Remote Sens. Ecol. Conservation*, vol. 8, no. 2, pp. 236–250, Apr. 2022.
- [9] K. S. P. Premarathna, R. M. K. T. Rathnayaka, and J. Charles, "An elephant detection system to prevent human-elephant conflict and tracking of elephant using deep learning," in *Proc. 5th Int. Conf. Inf. Technol. Res.*, Dec. 2020, pp. 11–13.
- [10] K. S. P. Premarathna and R. M. K. T. Rathnayaka, "CNN based image detection system for elephant directions to reduce human-elephant conflict," in *Proc. 13th Int. Res. Conf. Gen. Sir John Kotelawala Defence Univ.*, 2020, pp. 128–135.
- [11] S. Ravikumar, D. Vinod, G. Ramesh, S. R. Pulari, and S. Mathi, "A layered approach to detect elephants in live surveillance video streams using convolution neural networks," *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6291–6298, May 2020.
- [12] S. Pemasinghe and P. K. W. Abeygunawardhana, "Development of an elephant detection and repellent system based on EfficientDet-lite models," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Jan. 2023, pp. 1–6.

- [13] M. Jothibasu, M. Sowmiya, R. Harsha, K. Naveen, and T. B. Suriyaprakash, "Improvement of computer vision-based elephant intrusion detection system (EIDS) with deep learning models," *Innov. Eng. AI Appl.*, vol. 10, pp. 131–153, Jul. 2023.
- [14] M. K. Mondal, R. Mandal, S. Banerjee, U. Biswas, J. C.-W. Lin, O. Alfarraj, and A. Tolba, "Design and development of a fog-assisted elephant corridor over a railway track," *Sustainability*, vol. 15, no. 7, p. 5944, Mar. 2023.
- [15] V. Moorthy and V. Rukkumani, "Deep vision-based wildlife intrusion detection with colour distribution preserved generative adversarial networks," *Pattern Recognit.*, vol. 161, May 2025, Art. no. 111272.
- [16] S. R. M. Sudhanan, A. K. Rahuman, and S. M. M. Roomi, "Using mask R-CNN and distance transform to detect elephants on railway tracks for collision prevention," *Measurement*, vol. 248, May 2025, Art. no. 116951.
- [17] J. Kommalapati, S. B. Dasari, and H. Pidaparathi, "Drone-based elephant detection and counting with YOLOv5," in *Proc. 4th Int. Conf. Pervasive Comput. Social Netw. (ICPCSN)*, May 2024, pp. 527–532.
- [18] S. Sawal and S. S. Kathait, "MobileNetV2: Transfer learning for elephant detection," *Int. J. Comput. Appl.*, vol. 186, no. 61, pp. 59–65, Jan. 2025.
- [19] M. Zeppelzauer, A. S. Stöger, and C. Breiteneder, "Acoustic detection of elephant presence in noisy environments," in *Proc. 2nd ACM Int. Workshop Multimedia Anal. Ecol. Data*, Oct. 2013, pp. 3–8.
- [20] J. Björck, B. Rappazzo, D. Chen, R. J. Bernstein, P. H. Wrege, and C. P. Gomes, "Automatic detection and compression for passive acoustic monitoring of the African forest elephant," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 476–484.
- [21] J. N. Prince and S. J. Sugumar, "Surveillance and tracking of elephants using vocal spectral information," *Int. J. Res. Eng. Technol.*, vol. 3, no. 19, pp. 664–671, May 2014.
- [22] A. Singh and M. V. Ramesh, "Design and simulation of elephant intrusion detection system," in *Proc. 6th Int. Workshop Comput. Sci. Eng. (WCSE)*, Tokyo, Japan, 2016, pp. 749–753.
- [23] U. Ranasinghe, S. Abeyrathne, L. Samaranyake, T. Weerakoon, N. Harischandra, and G. Dissanayake, "Enhanced frequency domain analysis for detecting wild elephants in Asia using acoustics," in *Proc. IEEE 17th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Aug. 2023, pp. 140–145.
- [24] C. Ramasubramanian, S. Lokiah, Y. Viswanath, and S. Jamthe, "Averting human-elephant conflict using IoT and machine learning of elephant vocalizations," in *Proc. IEEE 8th World Forum Internet Things (WF-IoT)*, Oct. 2022, pp. 1–6.
- [25] M. E. Deowan, A. K. Nuhel, I. Haider, M. E. A. A. M. Tusher, M. M. Sazid, and M. N. M. Bhuiyan, "A warning and detection system for elephant intrusions utilizing signal processing and IoT," in *Proc. 4th Int. Conf. Electr., Comput. Telecommun. Eng. (ICECTE)*, Dec. 2022, pp. 1–4.
- [26] H. Dewmini, D. Meedeniya, and C. Perera, "Elephant sound classification using deep learning optimization," *Sensors*, vol. 25, no. 2, p. 352, Jan. 2025.
- [27] C. M. Geldenhuys and T. R. Niesler, "Learning to rumble: Automated elephant call and sub-call classification, detection and endpointing using deep architectures," *Bioacoustics*, vol. 34, no. 3, pp. 307–354, May 2025.
- [28] A. P. Sayakkara, N. Jayasuriya, T. Ranathunga, C. Suduwella, N. Vithanage, C. Keppitiyagama, K. De Zoysa, K. Hewage, and T. Voigt, "Eloc: Locating wild elephants using low-cost infrasonic detectors," in *Proc. 13th Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS)*, Jun. 2017, pp. 44–52.
- [29] P. Fernando, J. Jayewardene, T. Prasad, W. Hendavitharana, J. Pastorini, and S. Lanka, "Current status of Asian elephants in Sri Lanka," *Gajah*, vol. 35, pp. 93–103, Jan. 2011.
- [30] K. Nozawa and T. Shotake, "Genetic differentiation among local populations of Asian elephant," *J. Zoolog. Systematics Evol. Res.*, vol. 28, no. 1, pp. 40–47, Apr. 2009.
- [31] M. Lee, J. Byun, D. Kim, J. Choi, and M. Kim, "Improved modified energy ratio method using a multi-window approach for accurate arrival picking," *J. Appl. Geophys.*, vol. 139, pp. 117–130, Apr. 2017.
- [32] B. Madhu, M. A. Rahman, A. Mukherjee, M. Z. Islam, R. Roy, and L. E. Ali, "A comparative study of support vector machine and artificial neural network for option price prediction," *J. Comput. Commun.*, vol. 9, no. 5, pp. 78–91, 2021.
- [33] D. Anguita, A. Ghio, S. Pischietta, and S. Ridella, "A hardware-friendly support vector machine for embedded automotive applications," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 1360–1364.
- [34] Q. Zhou, X. Yao, C. Wang, J. Hu, P. Liu, and J. Lin, "Adaptive moving ground-target detection method based on seismic signal," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [35] Z. Hu, X. Zhao, Y. Guo, and C. Zhou, "Intrusion target classification based on seismic signals using parallel neural network with time/frequency dimension reduction," *IEEE Sensors J.*, vol. 24, no. 5, pp. 7055–7063, Mar. 2024.
- [36] S. Hettigoda, C. Jayaminda, U. Amarathunga, S. Thaha, M. Wijesundara, and J. Wijekoon, "A geophone based surveillance system using neural networks and IoT," in *Proc. 2nd Int. Conf. Advancements Comput. (ICAC)*, vol. 1, Shiraz Thaha, Mali, Dec. 2020, pp. 488–493.
- [37] S. Cyriac, B. M. Harsha, and Y. Woon Kim, "Seismic activity-based human intrusion detection using deep neural networks," in *Proc. 13th Int. Conf. Inf. Commun. Technol. Conver. (ICTC)*, Oct. 2022, pp. 130–135.
- [38] B. Mukhopadhyay, S. Anchal, and S. Kar, "Detection of an intruder and prediction of his state of motion by using seismic sensor," *IEEE Sensors J.*, vol. 18, no. 2, pp. 703–712, Jan. 2018.
- [39] P. Choudhary, N. Goel, and M. Saini, "A seismic sensor based human activity recognition framework using deep learning," in *Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2021, pp. 1–8.
- [40] S. Anchal, B. Mukhopadhyay, and S. Kar, "Predicting gender from footfalls using a seismic sensor," in *Proc. 9th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2017, pp. 47–54.
- [41] D. S. Parihar, R. Ghosh, A. Akula, S. Kumar, and H. K. Sardana, "Seismic signal analysis for the characterisation of elephant movements in a forest environment," *Ecolog. Informat.*, vol. 64, Sep. 2021, Art. no. 101329.
- [42] D. Goderik, A. Westlund, G. Zetterqvist, F. Gustafsson, and G. Hendeby, "Seismic detection of elephant footsteps," in *Proc. 27th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2024, pp. 1–8.
- [43] M. S. Nakandala, S. S. Namasivayam, D. P. Chandima, and L. Udawatta, "Detecting wild elephants via WSN for early warning system," in *Proc. 7th Int. Conf. Inf. Autom. Sustainability*, Dec. 2014, pp. 1–6.
- [44] P. P. S. Fernando, K. Y. L. Perera, P. N. Dissanayake, J. A. D. M. Jayakody, J. L. Wijekoon, and M. Wijesundara, "Gaja-mithuru: Smart elephant monitoring and tracking system," in *Proc. 11th IEEE Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Nov. 2020, pp. 461–467.
- [45] C. Thilakarathne, D. Samarawickrama, H. Serasinghe, N. Kuruwitaarachchi, and U. Samarantunga, "Prototype solution for human-elephant conflict using secondary data," in *Proc. IEEE 7th Int. Conf. Conver. Technol. (I2CT)*, Apr. 2022, pp. 1–6.
- [46] D. S. Parihar, R. Ghosh, A. Akula, S. Kumar, and H. K. Sardana, "Variational mode decomposition of seismic signals for detection of moving elephants," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022.
- [47] J. Vidunath, C. Shamal, R. Hiroshan, U. Gamlath, C. U. S. Edussooriya, and S. R. Munasinghe, "Identification of elephant rumbles in seismic infrasonic signals using spectrogram-based machine learning," *Appl. Syst. Innov.*, vol. 7, no. 6, p. 117, Nov. 2024.
- [48] C. Chandan, M. Chakraborty, S. Anchal, B. Mukhopadhyay, and S. Kar, "GajGamani: Mitigating man-animal conflict by detecting moving elephants using ground vibration-based seismic sensor," *IEEE Sensors Lett.*, vol. 8, pp. 1–4, 2024.
- [49] S. Deb, R. Ravindran, and S. K. Radhakrishnan, "Implementation strategy and performance analysis of a novel ground vibration-based elephant deterrent system," *J. Threatened Taxa*, vol. 17, no. 3, pp. 26704–26714, Mar. 2025.
- [50] D. S. Parihar and R. Ghosh, "A hybrid CNN-BiLSTM model for seismic signal based wildlife detection nearby railway tracks," *Authorea Preprints*, vol. 2025, pp. 1–16, Mar. 2025.
- [51] R. Steinmann, T. Nissen-Meyer, F. Cotton, F. Tilmann, and B. Mortimer, "Decoding the footsteps of the African savanna: Classifying wildlife using seismic signals and machine learning," *Methods Ecol. Evol.*, vol. 16, no. 4, pp. 819–836, Apr. 2025.
- [52] T. V. R. Vallabhan, N. Martin, R. Ramesh, J. Jaimes, V. Gautham, and R. Gouthamkrishna, "Seismic spectrogram based elephant detection system," in *Proc. 11th Int. Conf. Adv. Comput. Commun. (ICACC)*, Nov. 2024, pp. 1–6.
- [53] A. M. J. V. Costa, C. S. Pallikkonda, H. H. R. Hiroshan, G. R. U. Y. Gamlath, S. R. Munasinghe, and C. U. S. Edussooriya, "Detection of seismic infrasonic elephant rumbles using spectrogram-based machine learning," 2023, *arXiv:2312.02831*.

- [54] C. E. O'Connell-Rodwell, B. T. Arnason, and L. A. Hart, "Seismic properties of Asian elephant (*Elephas maximus*) vocalizations and locomotion," *J. Acoust. Soc. Amer.*, vol. 108, no. 6, pp. 3066–3072, Dec. 2000.
- [55] J. Akram and D. W. Eaton, "Adaptive microseismic event detection and automatic time picking," in *Proc. CSEG Annu. Conv.*, 2012.
- [56] A. Trnkoczy, "Understanding and parameter setting of STA/LTA trigger algorithm," in *New Manual of Seismological Observatory Practice (NMSOP)*. Deutsches GeoForschungsZentrum GFZ, 2012, pp. 1–20.
- [57] L. Han, J. Wong, and J. C. Bancroft, "Time picking and random noise reduction on microseismic data," *CREWES Res. Rep.*, vol. 21, pp. 1–13, Jan. 2009.
- [58] D. C. Toledo-Pérez, J. Rodríguez-Reséndiz, and R. A. Gómez-Loenzo, "A study of computing zero crossing methods and an improved proposal for EMG signals," *IEEE Access*, vol. 8, pp. 8783–8790, 2020.
- [59] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Princeton, NJ, USA: Citeaser, 2005, p. 452.
- [60] R. Tibi, C. Young, A. Gonzales, S. Ballard, and A. Encarnacao, "Rapid and robust cross-correlation-based seismic signal identification using an approximate nearest neighbor method," *Bull. Seismolog. Soc. Amer.*, pp. 1954–1968, Jul. 2017.
- [61] J. G. Proakis, *Digital Signal Processing: Principles, Algorithms, and Applications, 4/E*. London, U.K.: Pearson, 2007.
- [62] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
- [63] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. 3rd Int. Conf. Knowl. Discovery Data Mining*, 1994, pp. 359–370.
- [64] U. Kumar, C. P. Legendre, L. Zhao, and B. F. Chao, "Dynamic time warping as an alternative to windowed cross correlation in seismological applications," *Seismolog. Res. Lett.*, vol. 93, no. 3, pp. 1909–1921, May 2022.
- [65] P. Senin, "Dynamic time warping algorithm review," Dept. Inf. Comput. Sci., Univ. Hawaii at Manoa Honolulu, Honolulu, HI, USA, Tech. Rep., 2008, vol. 855, no. 1–23, p. 40.
- [66] S. McCrary, "Implementing algorithms to measure common statistics," *SSRN Electron. J.*, vol. 2015, Oct. 2015, Art. no. 2695198.
- [67] M. Norton, A. Mafusalov, and S. Uryasev, "Soft margin support vector classification as buffered probability minimization," *J. Mach. Learn. Res.*, vol. 18, no. 68, pp. 1–43, 2017.
- [68] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
- [69] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, vol. 2, 1995, pp. 1137–1143.
- [70] J. He, L. Li, J. Xu, and C. Zheng, "ReLU deep neural networks and linear finite elements," 2018, *arXiv:1807.03973*.
- [71] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4. Cham, Switzerland: Springer, 2006.
- [72] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
- [73] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
- [74] M. A. Nielsen, *Neural Networks and Deep Learning*, vol. 25. San Francisco, CA, USA: Determination press, 2015.

- [75] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.
- [76] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, no. 3, pp. 647–665, Dec. 2014.



JALIYA L. WIJAYARAJA (Member, IEEE) received the B.Sc. Engineering degree (Hons.) in electrical and electronic engineering from Sri Lanka Institute of Information Technology (SLIIT), Sri Lanka, in 2020, where he is currently pursuing the M.Phil. degree in computer networks.

Since 2021, he has been a Research Assistant at the Department of Computer Systems Engineering, SLIIT. His research interests include electronic design, signal processing, the IoT systems.



JANAKA L. WIJEKON (Senior Member, IEEE) received the B.Sc. degree from SLIIT, Sri Lanka, in 2010, and the M.Sc. and Ph.D. degrees from Keio University, Japan, in 2013 and 2016, respectively. With a background as a Postdoctoral Researcher with extensive experience in academia, his expertise centres around leveraging the AIoT to advance smart societies. His research interests include collaborative smart infrastructural design and implementations for secure and efficient resource allocation in smart societies and protocol design for data exchange among stakeholders to improve agricultural productivity.



MALITHA WIJESUNDARA (Member, IEEE) received the B.Eng. degree (Hons.) in electronic engineering from the University of Warwick, U.K., in 1998, and the Ph.D. degree in computer engineering from the National University of Singapore. He has extensive experience in embedded systems, wireless sensor networks, media streaming for e-learning, and energy harvesting for elephant tracking. His research interests include developing technology solutions to challenges faced by rural and underprivileged communities. He is a member of ISACA and CS (SL).

...