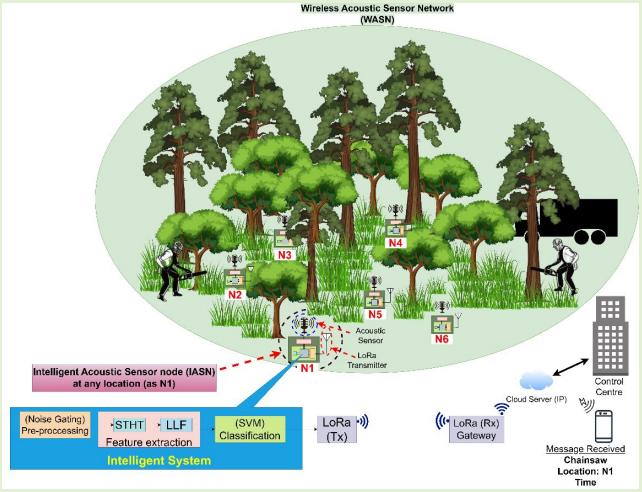
Real-Time Monitoring of Illegal Logging Events Using Intelligent Acoustic Sensors Nodes

Vivek Singh, *Student Member, IEEE*, Kailash Chandra Ray, *Member, IEEE*, and Somanath Tripathy, *Senior Member, IEEE*

***Abstract*—The usage of sensor networks for monitoring and assessing various facets of our everyday life has grown significantly. Illegal logging activity identification is critical in today's forest environment to conserve habitat. Acoustic signals are a significant tool in monitoring illegal logging activities over wireless acoustic sensor networks (WASNs). However, the continuous monitoring of acoustic signals requires a high computational demand that prevents real-time computations within the nodes. This article presents an efficient methodology based on acoustic signal processing and its classification for real-time remote monitoring and detection of illegal logging activities. The proposed method consists of a short-time Hartley transform (STHT)-based spectrogram for extracting the low-level audio features (LLFs) from the acoustic signals. Subsequently, these extracted features are classified into five acoustic classes using k-nearest neighbor (kNN), decision tree (DT), random forest (RF), adaptive boosting (Ada-Boost), and support vector machines (SVMs) classifiers. The efficient combination of feature representation and classification method is implemented on a 32-bit microprocessor platform, interfaced with a LoRa wireless module to develop an intelligent acoustic sensor nodes (IASN) system, which is validated by regenerating various acoustic signals along with forest ambience using loudspeakers in an experimental laboratory setup. Once an illegal logging activity is detected, the edge node generates an alert message and sends it to the mobile phone through the gateway. The accuracy of the proposed system is 96.61%, along with sensitivity, specificity, and *F*-score of 96.20%, 99.14%, and 96.07%, respectively.**



***Index Terms*—Illegal logging monitoring, intelligent acoustic sensor nodes (IASN), LoRa, short-time Hartley transform (STHT), support vector machine (SVM).**

I. Introduction

A. Context of the Study

FOREST has an imperative role in maintaining the earth's global biodiversity and preserving the ecological balance. The land cover under the forest is estimated at 4.06 billion hectares [1], which is 32.11% of the earth's total land area [2]. It supports almost 90% of the terrestrial biodiversity [3] and plays an important role in preventing soil erosion and landslides [4]. However, annually, average global forest loss is estimated at around 25 million hectares [5], and its dominant driver is illegal logging [5] activities. Illegal logging can cause unmanageable and irreparable deforestation. To check such activities, the existing forest monitoring systems rely mainly upon ground staff patrolling [3], [7], which is very expensive, time-consuming and requires a large number of workforce resources. In forest scenarios, the technology-based solutions, including the acoustic sensors, can play a crucial role in automating remote monitoring and detecting illegal logging in real time [6]. As tree cutting produces significant and distinct sound, an acoustic-based, that is, a wireless acoustic sensor network (WASN) can serve as the most feasible solution. To a certain extent, the research community has shown the effectiveness of event detection systems [30], [41], [42], [43], [44] using acoustic sensors. However, a complete solution is still lacking due to factors such as 1) the increased demand to process vast amounts of acoustic data; 2) emerging IoT-based technologies for remote forest monitoring; and 3)

Manuscript received 10 January 2024; revised 13 June 2024; accepted 21 June 2024. Date of publication 17 July 2024; date of current version 1 September 2024. This work was supported in part by the Ministry of Electronics and Information Technology (MeitY), Government of India, New Delhi, through the project "Special Manpower Development Program for Chips to System Design," under Project R&D/SP/EE/DEIT/SMD/201516/126; and in part by the Visvesvaraya Ph.D. Scheme. The associate editor coordinating the review of this article and approving it for publication was Dr. Lin Wang. *(Corresponding author: Vivek Singh.)*

Vivek Singh and Kailash Chandra Ray are with the Department of Electrical Engineering, Indian Institute of Technology Patna, Bihta, Patna 801106, India (e-mail: vivek.pee16@iitp.ac.in; kcr@iitp.ac.in).

Somanath Tripathy is with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, Bihta, Patna 801106, India (e-mail: som@iitp.ac.in).

Digital Object Identifier 10.1109/JSEN.2024.3419897

the necessity to develop low-cost, low-energy consuming, high-speed, and long-life devices [9]. Acoustic signal processing and machine learning techniques are required to automate the process, which can be employed in the WASN system in three configurations [8]. First, both acoustic feature extraction and classification tasks are implemented on a cloud server [8]. Second, the feature extraction task is implemented on the remote edge device and the classification task on the cloud server [8], and third, acoustic feature extraction and classification tasks are implemented on the remote edge device [8]. For acoustic-based remote monitoring applications, the input signal is continuous for an infinite time; thus, first and second configurations require continuously large data transmission using wireless communication modules. However, the power requirement of radio modules generally has a higher value than the power requirement of edge computing devices [8]. The third configuration is most suited to increase the battery-powered edge node's lifespan, subject to an energy-efficient algorithm for acoustic feature extraction and classification for remote monitoring of illegal logging activities.

B. Related Works

Several studies have reported the remote monitoring of illegal logging activities using acoustic sensors. The recorded data using the acoustic sensors is processed using a combination of feature extraction and classification algorithms [10], [11], [12], [13], [14], [15], [16], [46], [47], [48], [49]. Among the various feature extraction methods reported, the time-domain features such as temporal magnitude features [6], [10], [14], TESPAR features [17], [18], and auto-correlation features [17] have low computational complexity. These time-domain features fail to provide the frequency information of the acoustic signal. The frequency-related features, namely frequency magnitude [14], [19], and Haar features [20], are extracted by employing frequency domain tools such as Fourier transform (FT). However, frequency domain features lack time-domain information, which is overcome by time–frequency domain features as in [21], [22], [23] and [24]. The low-level acoustic features [22], [23], [42], [43], [44] and spectrogram features [21] are extracted from the time–frequency representation of the acoustic signals. The time–frequency domain features also have a limitation in that they fail to address the dynamic range of amplitudes of acoustic signals as processed in ears targeting the practical scenario. To address this issue, the Mel-scaled features, which are biologically inspired, such as Mel-frequency cepstral coefficients (MFCC) as in [25], [26] and [27] are combined with classifiers such as distance measure [6], [17], [18], [22], thresholding [10], [14], [20], Gaussian mixture model (GMM) [18], spectral feature-based Gauss-Bayesian classifier (SGBC) [25], decision tree (DT) [17], support vector machine (SVM) [18], [23], [27], [28], [29], deep neural network (DNN) [30], and a convolutional neural network (CNN) [21] for automated classification of illegal logging activities.

C. Research Gap

In each of the above-reported works reported earlier, a number of limitations have been observed. In [10], [11], [12], [13] and [14], the use of a fixed threshold-based technique is implemented, which lacks reliability for different scenarios of the forest environment and hence, variable thresholds must be used for accurate detection. Further, the background noise is also associated with the input acoustic signal; therefore, the appropriate choice of threshold is a challenge failing, which results in the generation of a higher number of false positives. In [10], [11] and [12], the wireless technology used is limited for low-range data transmission; these have limitations to deploy in large forest areas. In [15], [17], [18], and [19] the authors only mentioned the alert system conceptually without any suitable hardware implementations. In most of the works, such as [10], [12], [21] and [27], the number of illegal classes considered is only chainsaw, and no other illegal activities are considered. In addition, the resolution of the reported works is low, that is, the actual length of acoustic signal employed for illegal logging detection is at least a minimum of 3 s in duration. All of these limitations mentioned above are addressed in this study to develop efficient monitoring of illegal activities in forests.

D. Motivation, Objective, and Contribution

The feature extraction and classification steps are significant in processing and monitoring the acoustic sensors data; thus, the efficient utilization of hardware resources greatly depends on these. The lightweight techniques with lower computational complexity are required in contrast to the methodologies reported in the literature. The methodology proposed in this study is based on a short-time Hartley transform (STHT), which is deemed to have low computational complexity and extract low-level features from the pre-processed acoustic sensor data. The extracted acoustic features are classified using the k-nearest neighbor (kNN), DT, random forest (RF), adaptive boosting (Ada-Boost), and SVMs classifiers. The efficient method, that is, a combination of feature extraction and classification method, reported the higher accuracy and is then implemented on a 64-bit microprocessor platform interfaced with a LoRa module to develop an intelligent acoustic sensor nodes (IASN) system. The developed IASN system is validated over the recorded data from acoustic sensors capable of detecting into one of five acoustic classes, that is, chainsaw, handsaw, vehicle, speech, and forest ambience. The IASN system will serve as an assistive remote forest monitoring system, thus reducing the time required to prevent illegal logging. The contribution of this study is twofold:

1) To propose an efficient algorithm with low computational load.
2) To develop an IASN system with necessary facilities for seamless remote monitoring and detection of illegal logging activities.

The rest of the article is organized as follows. Section II presents the theory of existing signal decomposition and classification methods. The proposed method for detecting illegal logging sounds is explained in Section III. Section IV presents the hardware system architecture and laboratory experiment. The obtained results are discussed and analyzed in Section V. Finally, this study is concluded in Section VI.

II. Background of STHT

The Hartley transform (HT) [31] maps a time-domain real-valued function into a frequency domain real-valued function. As the discrete Hartley transform (DHT) computes only real-valued coefficients in contrast to complex-valued coefficients computed in the discrete FT, the computational complexity of signal decomposition is reduced by fourfold. One complex multiplication requires four real multiplications and two additions. Given a discrete function $x[n]$, its DHT [32] is as follows:

$$H[k] = N^{-1} \sum_{n=0}^{N-1} x[n] \text{cas}\left(\frac{2\pi kn}{N}\right) \tag{1}$$

where $\text{cas}(2\pi kn/N) = \cos(2\pi kn/N) + \sin(2\pi kn/N)$. The corresponding inverse discrete Hartley transform (IDHT) [32] can be written as follows:

$$x[n] = \sum_{k=0}^{N-1} H[k] \text{cas}\left(\frac{2\pi kn}{N}\right). \tag{2}$$

Another form of HT, that is, discrete STHT, is an obvious choice to analyze the time-varying real signal. Hence, discrete STHT [33] is employed to develop the proposed method for feature extraction of acoustic events in the forest. The discrete STHT of non-stationary signal $x[n]$ is obtained as

$$H_m[k] = \sum_{n=-\infty}^{\infty} w[m-n] x[n] \text{cas}\left(\frac{2\pi kn}{N}\right) \tag{3}$$

where $H_m[k]$ is a 2-D vector, one dimension being the time instants, that is, m and another being a frequency bin, that is, k. $H_m[k]$ is obtained at each time sample m by weighting the sequence $x[n]$ by the window $w[m-n]$. In practice, the limits of (3) are finite and are determined by the length of $w[m]$. Bracewell in [34] introduced the fast Hartley transform (FHT) algorithm to compute HT with minimum hardware complexity.

III. Proposed Methodology

In this study, a novel method consists of three significant steps, that is, pre-processing using noise gating technique, feature extraction using STHT-based low-level audio features (LLFs), and classification using SVM is employed. First, the audio dataset is prepared from the sounds of different activities recorded in the forest and the data collected from open-source internet repositories. Next, the noise gating technique is employed to acquire the ambience noise signature and filter the input audio signal using it. This filtered audio signal is fed into the feature extraction unit to extract the proposed STHT-based LLFs. The SVM classifier classifies these extracted features into one of the five audio classes. The workflow of the proposed method is depicted in Fig. 1.

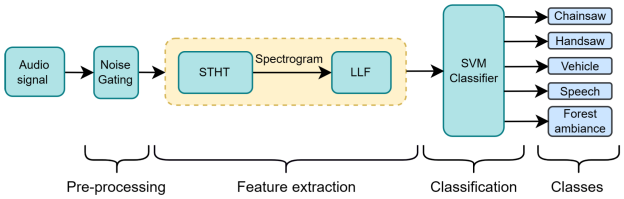


Fig. 1. Block diagram of the proposed method.

A. Data Preparation

This study considers five categories of audio classes such as chainsaw, handsaw, vehicle, voice, and forest ambience. However, to the authors' knowledge, the standard dataset for the problem addressed in this study is not readily available. Therefore, the dataset for the experiment is prepared by recording the audio of the considered audio classes in the forest regions of the states of Odisha and Uttar Pradesh, India, in real time. The recording setup consists of an electret unidirectional microphone, AHUJA ASM-780XLR, along with an audio recorder, Zoom H1n. The recording is made at a 48 000-Hz sampling rate with a 16-bit resolution. Along with these recordings, audio samples have also been collected from open-source libraries, such as *freesounds.org* [36], [37], [38], [39] and Google's *Audioset* [40]. It is worth mentioning that the collected dataset has large variability concerning the type of device used for recording, location, surrounding noises, and type of equipment/tool/engine. All the collected data, including recordings made in the forest and the data collected from internet resources, are regenerated at a sampling rate of 16 kHz with 16-bit resolution and are randomly shuffled to obtain an accumulated dataset. The individual files in this accumulated dataset are then clipped into 1-s audio segments, resulting in a total number of 64 833 clips (where chainsaw: 21 774, handsaw: 3753, vehicle: 10 267, speech: 11 667, and forest ambience: 17 372). To train and validate the system, an audio augmentation technique has been employed on these individual clips to enlarge the dataset size and have additional variability in the audio samples. The adopted audio augmentation techniques are described here briefly:

1) *Pitch Shifting:* It is the rising or lowering of the audio clip's pitch randomly in the pitch range $\alpha(t)$ subject to an audio class, $a(t) = \Sigma_i x_i(t) \exp(j\phi_i(t))$, where $\phi_i(t) = \Sigma_t \alpha(t)\omega_i(t)$ is an instantaneous phase, $\omega_i(t)$ is the instantaneous frequency, A_i is the amplitude, and $a(t)$ is the augmented clip.
2) *Time Reversal:* The audio clip is flipped along the time axis, $a(t) = x(-t)$.
3) *Forward and Backward Shifting:* The audio clip is randomly shifted either forward or backwards with rollover along the time axis. The amount of shifting is chosen randomly, $a(t) = x(t - T)$.
4) *Polarity Inversion:* Amplitude at each time instant is multiplied with -1, $a(t) = -x(t)$.
5) *Background Noise:* Forest ambience is considered as a background noise. Each augment $a(t)$ is obtained using $a(t) = (1-w)\cdot x(t) + w \cdot g(t)$, where $x(t)$ is the original audio signal, $g(t)$ is the background signal, and w is a weighting value generated at random for each augment.

The employment of an audio augmentation technique resulted in a corpus of 388 993 one-second audio clips, which includes all considered five classes. This audio clip's corpus is then randomly split into training and testing sets in a 70%–30% ratio. Table I summarizes the training and testing dataset with the class-wise distribution. The training

TABLE I
SUMMARY OF DATASET CREATED USING AUDIO AUGMENTATION TECHNIQUE

Audio Class	No. of files (One second each) Total	Training	Testing
Chainsaw - C	130641	91449	39192
Handsaw - H	22519	15764	6755
Vehicle - V	61600	43120	18480
Speech - S	70000	49000	21000
Forest ambience - F	104233	72964	31269
Total	388993	272297	116696

dataset is used to train the SVM classifier on the proposed audio features, whereas the testing set is used to evaluate the performance of the classifier when implemented on the COTS components-based developed hardware platform in the laboratory experimental setup. It is to be noted that ten-fold cross-validation is performed on the entire dataset to evaluate the performance of the classifier model.

B. Pre-Processing: Noise Gating

The raw audio signal is associated with different kinds of noises in real time. Removing noise from audio signals is crucial in correctly detecting different audio classes. The proposed method adopts the noise gating technique [41] for ambience noise reduction, as depicted in Fig. 2. Since the typical acoustic environment lasts around 1 h, we have updated the noise signal as 1 h passes by recording the fresh 120-s environmental noise signal and computing its frequency content. While performing the experiments in a controlled laboratory environment, we ensured, for these 120 s, that no acoustic events of interest were available in the acoustic environment. This frequency content, that is, the magnitude of each frequency bin available for this 120-s, is computed using DTH.

In this technique, the audio noise signature of an acoustic ambience is established for filtering. At first, the ambience noise signal $s[n]$ (i.e., when the audio signal of interest is not present in the acoustic ambience) is recorded and transformed into frequency domain $S[k]$ using DHT given in (4). The magnitude of each frequency bin of the noise signal is accepted as the signature of the ambient noise. This noise signature serves as a gate for filtering the noisy audio signal $x_s[n]$. It is to be noted that the ambience noise is dynamic in nature and varies with time; thus, the noise signature is updated hourly in laboratory experiments to address it

$$\begin{aligned}\mathrm{DHT}[s[n]] &= S[k] \\ &= \frac{1}{N}\sum_{n=0}^{N-1} s[n]\left(\sin\left(\frac{2\pi kn}{N}\right)+\cos\left(\frac{2\pi kn}{N}\right)\right)\end{aligned} \tag{4}$$

$$X_s[k] = \frac{1}{N}\sum_{n=0}^{N-1} x_s[n]\left(\sin\left(\frac{2\pi kn}{N}\right)+\cos\left(\frac{2\pi kn}{N}\right)\right). \tag{5}$$

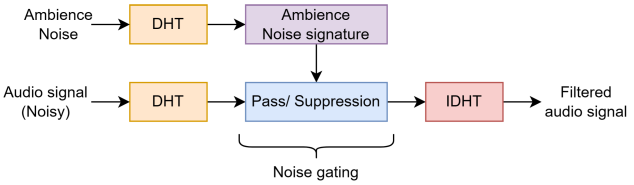


Fig. 2. Filtering using noise gating technique.

Next, the recorded noisy audio signal, $x_s[n]$, is transformed into the frequency domain $X_s[k]$ using DHT as in (5). The ambience noise from this noisy audio signal is removed in the frequency domain by comparing the magnitude of each frequency bin of $X_s[k]$ to the noise signature, $S[k]$, obtained in (4). The comparison is made as described by (6), that is, if the magnitude of the frequency bin in the input noisy audio signal ($x_s[n]$) is greater than the magnitude of the corresponding bin in the noise signature, the frequencies corresponding to that bin are passed through to the output. However, if the magnitude of the bin in the input noisy signal is equal to or below the corresponding bin in the noise signature, it is attenuated by a factor g, which is set at -10 dB in this study to improve the SNR. This filtering process is represented as follows:

$$X[k] = \begin{cases} X_s[k], & \text{if } X_s[k] > S[k] \\ gX_s[k], & \text{if } X_s[k] \le S[k] \end{cases} \tag{6}$$

where $k = 0, 1, \ldots, N-1$ is the bin number for N-sample points transformed to the frequency domain. The filtered signal is subsequently computed using an IDHT as given as follows:

$$x[n] = \sum_{k=0}^{N-1} X[k]\left(\sin\left(\frac{2\pi kn}{N}\right)+\cos\left(\frac{2\pi kn}{N}\right)\right). \tag{7}$$

This technique behaves as a multiple bandpass or band-reject filtering. The filtered audio signal is fed into the feature extraction stage to extract the proposed LLFs.

C. Proposed Feature Extraction

The filtered audio signal from the noise gating unit is fetched into the proposed feature extraction unit. The raw audio signal is a non-stationary time-domain signal, which needs to decompose to extract useful features with reduced dimensionality from it. The proposed feature extraction employs the STHT, (3), discussed under Section II, to transform the time-domain audio signal into a time–frequency domain. STHT uses real-valued mathematics to decompose the time-domain audio signal into a real-valued time–frequency domain function. Real-valued mathematics has less computational load than complex-value mathematics, as mentioned in Section II. In general, signal decomposition techniques such as the FT use complex-value mathematics to extract the phase information from the signal. However, the proposed features do not require the phase information of the audio signal. The computational complexity of the conventional short-time FT for N samples is $N.\log N$. However, the complexity of DHT is very much reduced by more than half; as such, it requires 16 624 operations in terms of additions and multiplications for $N = 1024$.

TABLE II
DESCRIPTION OF LOW-LEVEL FEATURES EXTRACTED FROM STHT-BASED SPECTROGRAM

Feature Name	Definition	Equation	Feature Name	Definition	Equation
Spectral centroid	Spectral centroid μ_1, indicates where the centre of mass of the spectrum is located.	$\mu_1 = \frac{\sum_{k=0}^{K} f_k S_k}{\sum_{k=0}^{K} S_k}$	**Bandwidth**	Spectral bandwidth is the second-order statistical value determining the low-bandwidth sound from the high-frequency sound [26].	$\sqrt{\frac{\sum_{k=0}^{K}(k-\mu_1)^2 (S_k)^2}{\sum_{k=0}^{K}(S_k)^2}}$
Spectral centre	It is the measure of the median frequency present in the signal spectrum. It balances the higher and lower energies.	$median(S_k)$	**Spectral flatness**	Spectral flatness is the measure of uniformity in the frequency distribution of the power spectrum. It distinguishes between harmonic and noise-like sounds.	$\frac{G_{mean}(S_k)}{A_{mean}(S_k)}$
Spectral roll-off	The spectral roll-off point is the frequency, so 95% of the signal energy is contained below this frequency.	$roll-off\ point = i$ $such\ that,$ $\sum_{k=0}^{i} S_k = 0.95 \sum_{k=0}^{K} S_k$	**Spectral crest factor**	It determines how peaked is the power spectrum of the sound signal. It is higher for harmonic/ tonal sounds and lower for noise-like sounds.	$\frac{max(S_k)}{rms(S_k)}$
Spectral spread	It is described as the average deviation of the rate map around the centroid. This feature is closely related to the bandwidth of the sign [26].	$\mu_2 = \sqrt{\frac{\sum_{k=0}^{K}(f_k-\mu_1)^2 S_k}{\sum_{k=0}^{K} S_k}}$	**Entropy**	It is a measure of uniformity of flatness and is computed as Shannon's entropy.	$Sum(P_i\ log(P_i))$
Spectral skewness	Spectral skewness is the 3rd-order statistical value, and it measures the symmetry of the spectrum around its arithmetic mean value.	$\frac{\sum_{k=0}^{K}(f_k-\mu_1)^3 S_k}{(\mu_2)^3 \sum_{k=0}^{K} S_k}$	**Spectral flux**	The spectral flux is the 2-norm of the frame-to-frame spectral amplitude difference vector. It points to the sudden changes in the frequency energy distribution of the sound.	$\sqrt{\sum_{k=1}^{K}\{abs(S_k - S_{k-1})\}^2}$
Spectral kurtosis	Kurtosis is the 4th order statistical measure and describes the flatness of the spectrum around the mean value.	$\frac{\sum_{k=0}^{K}(f_k-\mu_1)^4 S_k}{(\mu_2)^4 \sum_{k=0}^{K} S_k}$	**OBSC**	Octave-based spectral contrast (OBSC) is the difference between peaks and valleys measured in sub-bands by octave scale filters [26].	-
Spectral slope	It is the measure of the slope of the amplitude of the signal and is computed by linear regression.	$\frac{\sum_{k=0}^{K}(f_k-\mu_f)(S_k-\mu_s)}{\sum_{k=0}^{K}(f_k-\mu_f)^2}$	**Standard deviation**	It is a measure of how much the magnitude of the spectrum has deviated from its mean value.	$\sqrt{\frac{\sum_{k=0}^{K}(S_k-\bar{S})}{K}}$
Spectral decrease	It measures the average spectral slope of the rate-map representation, putting a strong emphasis on low frequencies [26].	$\frac{\sum_{k=b+1}^{K}\{(S_k-S_b)/(k-1)\}}{\sum_{k=b+1}^{K} S_k}$	**Energy**	Total energy content in the frame. It is measured as root mean square.	$\sqrt{\frac{\sum_{k=0}^{K}(S_k)^2}{K}}$

f_k : frequency corresponding to bin k, S_k : spectral magnitude at bin k, K: total number of frequency bins, μ_s: mean spectral magnitude, μ_f: mean frequency, G_{mean}: geometric mean, A_{mean}: arithmetic mean, max: maximum, rms: root mean square, P_i: probability of sample class i, $\bar{S}$: mean value of the spectrum

The non-stationary property of the audio signal is addressed by decomposing it in small overlapping time frames of 32 ms, that is, 512 sample points with a 512-point FHT. The frames from the audio signal are windowed using the Hamming function while the overlapping between the adjacent frames is kept at 50%. This output of the STHT can be represented in 2-D form, that is, the spectrogram, as depicted in Fig. 3, which in turn represents the strength of different frequencies (bins) at a certain time frame. The resultant shape of the spectrogram obtained after performing STHT on1-s sound clip is 512×63. Where 512 is the number of frequency bins, and 63 is the number of time frames. This spectrogram output of the STHT is used for extracting sixteen LLF [26] summarized in Table II. The considered LLFs represent the physical and statistical properties of the audio signal. For each audio class, these properties are unique and thus have distinctive LLFs. For each frame, the extracted LLF resulted in 27 features per frame. The feature vector for a 1-s audio clip is obtained by concatenating these 27 features of each frame and flattening it out into a 1-D vector along the horizontal axis. The resulting feature vector length is 1701 coefficients, described in Table III. These extracted LLFs are provided as input to the classifier unit.

D. Classification

The proposed audio features extracted in the previous stage are classified into one of the considered five audio classes using a classifier. To select the most suitable classifier from the available ones, performance in terms of classification accuracy is compared between machine learning algorithms such as kNN, DT, RF, Ada-Boost, and SVM is conducted. The feature vector is pre-processed to have a mean value of zero and a standard deviation of one, that is, standardized before being provided as input to the classifier. The considered machine learning algorithms are trained and evaluated in the simulation environment using the proposed features extracted from the training and testing datasets, respectively. In this study, ten-fold cross-validation is performed on the entire dataset to evaluate the performance of the classifier model. The implementation of each of the classifier techniques over the features extracted using STHT-based low-level features is described below.

1) k-Nearest Neighbor: The kNN classifier is the simplest supervised learning method that depends upon the closest training patterns of the class label in a feature space. The test pattern is compared with all the patterns available for training to reach the closest pattern. It consists of two steps: the estimation of the nearest neighbors and the second is the identification of the class related to those neighbors. The test pattern is assigned to the class based on the majority votes of neighbors. In this study, the Manhattan distance metric has been considered for computing the distance between the test and training patterns. The value of k, that is, the number of nearest neighbors used, is 5. The overall classification accuracy achieved on the test dataset is 88.84%.

2) Decision Tree: The DT utilizes a tree-like structure where each leaf node represents the target, and each internal node represents the feature for solving problems. To implement it, the Gini index is used as the splitting criterion. The maximum

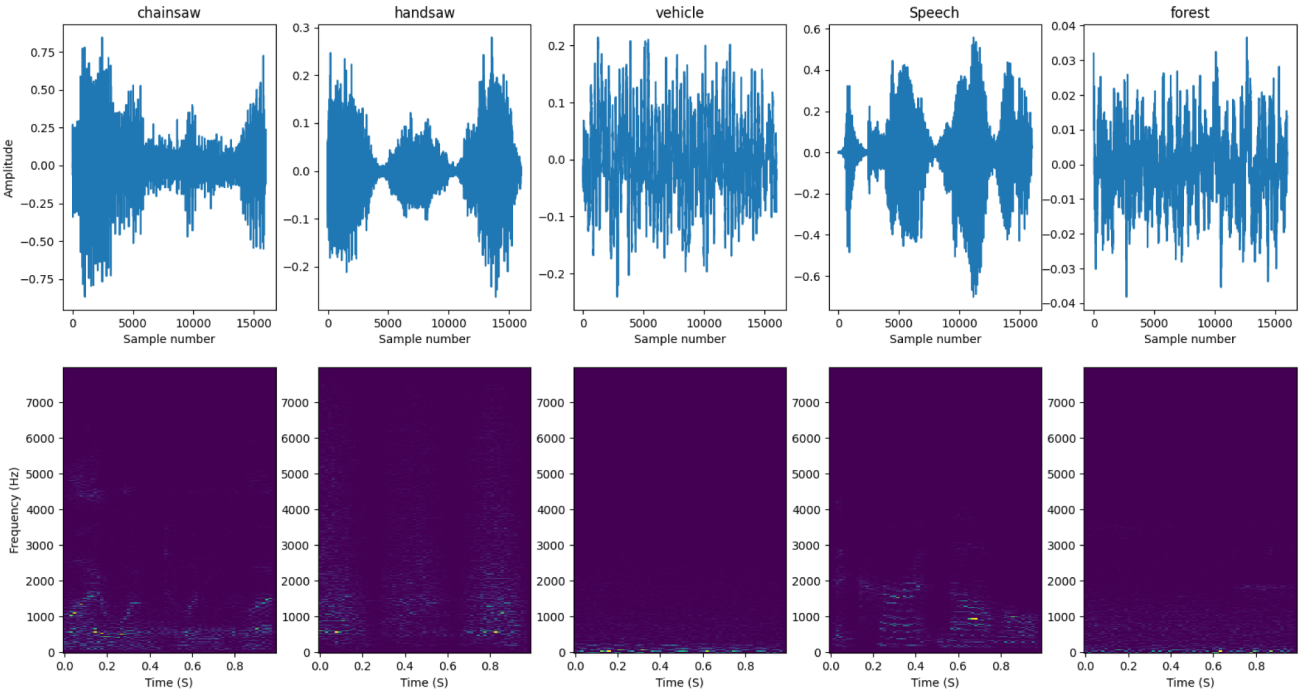


Fig. 3. Spectrogram of considered audio classes obtained using STHT.

TABLE III
EXTRACTED FEATURE VECTOR DESCRIPTION

Feature Index	Feature description
1-63	Spectral Centroid
64-126	Spectral Centre
127-189	Spectral Roll-off
190-252	Spectral Spread
253-315	Spectral Skewness
316-378	Spectral Kurtosis
379-441	Spectral Slope
442-504	Spectral Decrease
505-567	Spectral Bandwidth
568-630	Spectral Flatness
631-693	Spectral Crest Factor
694-756	Entropy
757-819	Spectral Flux
820-1575	Octave Based Spectral Contrast (OBSC)
1576-1638	Standard deviation
1639-1701	Energy

depth is set at 20, with each branch having a minimum of two splits. The overall accuracy achieved by the DT classifier is 87.16% on the test dataset.

3) Random Forest: An RF is an ensemble of tree-structured classifiers [11]. Every tree in the forest gives a unit vote, assigning each input to the most probable class label. The majority vote chooses the final outcome/classification. The maximum number of estimators is set at 100. The tree can grow up to its maximum depth depending on the classification accuracy. The RF classifier achieved a classification accuracy of 95.81% on the test dataset.

4) Adaptive Boosting: The Ada-Boost is an ensemble classifier in which DT classifiers are sequentially ensembled to form a new classifier that performs better than the DT itself [12]. In tree boosting, output from each DT is added sequentially

TABLE IV
CLASSIFICATION ACCURACY COMPARISON OF DIFFERENT CLASSIFIERS

Class	kNN	DT	RF	Ada-Boost	SVM
Chainsaw	96.43	91.00	97.32	97.29	98.00
Handsaw	89.29	70.46	87.67	81.54	93.94
Vehicle	98.71	88.22	94.92	95.30	96.10
Speech	57.23	84.93	97.97	96.90	97.83
Forest ambience	94.54	88.28	95.48	94.96	95.16
Overall	88.84	87.16	95.81	95.05	96.61

All values are in per cent.

such that each tree tries to reduce the errors of the previous tree. In short, boosting combines many weak learners in sequence to obtain a robust, efficient, and accurate learner. Fifty DTs are used with a learning rate of 1. The maximum depth of each tree is set at 20. The minimum slip of each branch is set at 2. The overall classification accuracy achieved on the test dataset is 95.05%.

5) Support Vector Machine: SVM [13] maps the non-separable data to the higher dimensional feature space using kernel trick and makes data easily separable. The SVM works on maximizing the margin by creating a hyperplane between the different classes. It provides good performance because its learning capacity does not depend on the dimensionality of the feature space of the data [14]. This study considers the radial basis function (RBF) kernel. The SVM classifier is trained in a one-versus-one approach. The penalty variable C is set at 10, and the gamma value is auto-tuned. The classification accuracy achieved by the SVM classifier is 96.61% on the test dataset.

Table IV summarizes the overall classification accuracy and class-wise classification accuracy achieved by each of the above-discussed classifiers. It can be observed that the SVM

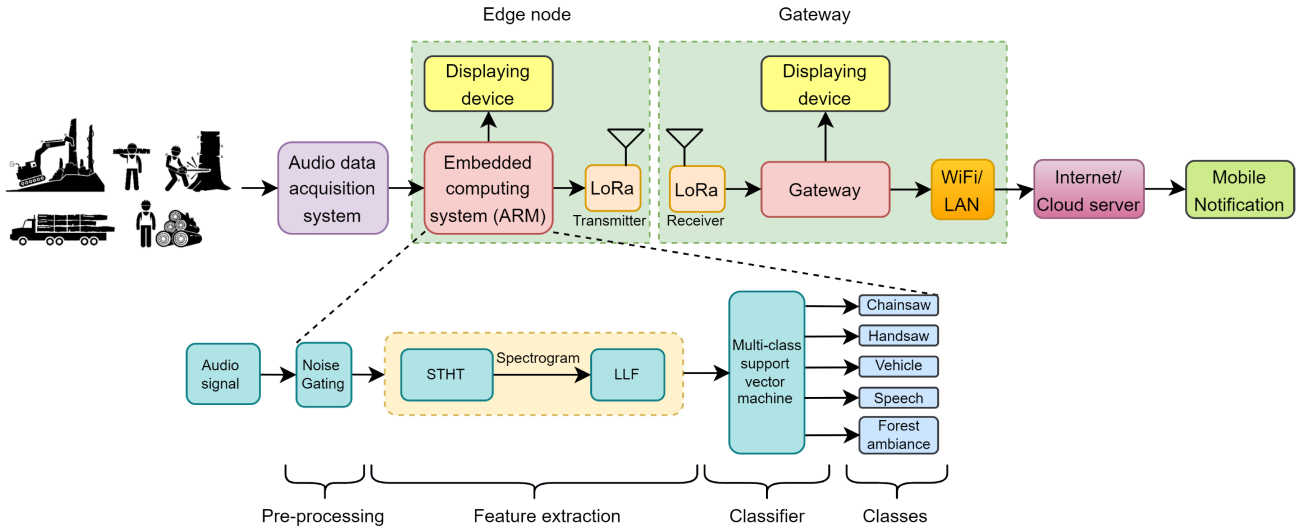


Fig. 4. Hardware system architecture for the proposed method.

classifier achieves the highest overall classification accuracy, that is, 96.61%. Also, the individual classes' classification accuracy of SVM is above 93%. Henceforth, the methodology, that is, the STHT-based LLFs with SVM classifier, is implemented on the developed hardware platform and discussed in detail.

Therefore, the proposed method is considered for further hardware implementation for the development of an efficient WASN system for remote monitoring of illegal logging activities.

IV. Hardware System Architecture and Implementation of Proposed Method

In this current study, two types of audio data are considered: 1) online database and 2) data collected in a dense forest region of Chitabeda, Odisha, India, using a microphone (AHUJA ASM-780XLR). The experiments are performed in a laboratory experimental environment by regenerating the audio signals using the speaker (ANKER A7910 with frequency response ranging from 70 to 20 kHz, and operating range of 20 m/66 ft). Thereafter, the audio signals are captured by the electret microphone and are taken to the 16-bit, 16 kHz ADC, which is further processed by the 32-bit microprocessor of the Raspberry Pi embedded platform using the proposed methodology.

A. IASN Architecture

The proposed method is realized on an embedded platform for real-time analysis and detection of illegal logging activities in laboratory experiments. The embedded hardware system architecture that represents IASN and employs the proposed methodology is depicted in Fig. 4. The WASN prototype is developed using commercially available off-the-shelf (COTS) components. It has six main subsystems: 1) the audio acquisition unit, which consists of an electret microphone (AHUJA ASM-780XLR with frequency response ranging from 50 to 16 000 Hz and sensitivity of 2 mV/Pa) and USB sound card (CM108). The quantization resolution is set at 16 bits, and the sampling frequency is 16 kHz. This unit converts the incoming analog audio signal to a digital audio signal. The sampled data is represented in a 32-bit floating-point format; 2) the processing unit is realized using a 32-bit ARM microprocessor-based embedded platform, that is, Raspberry Pi 3 Model B. It processes the acquired audio data using the proposed methodology in which the raw audio signal is pre-processed to reduce the background noises using the noise gating technique wherein, improving the signal-to-noise ratio (SNR). The proposed STHT-based LLFs are extracted and provided as input to the trained SVM classifier model to identify input signals into five subsequent classes. The detected audio signal class is displayed on the 16 × 2 LCD displaying module interfaced with the embedded platform; 3) the sensor node transmits the information using the data transmission unit, which uses Microchips' LoRa (line-of-sight distance of up to 15 km) module. It operates in the ISM band at 868 MHz; 4) the data transmitted by WASN is received at the LoRa-to-IP network gateway unit; 5) it re-transmits the received information to the cloud application server, that is, The Things Network; and 6) the mobile device is the end user, which fetches transmitted information from the cloud application as

TABLE V
Summary of Hyperparameters for Hardware Prototype

Parameter	Value
Sampling frequency	16000 Hz
Bit resolution	16-bits
Frame length	512 samples (32 ms)
Hop length	256 samples (16 ms)
Window	Hamming
No. of frames (N)	63
No. of frequency bins	512
Dimension of spectrogram	(512×63)
Feature vector length	1701 coefficients
SVM:	
Kernel function	Radial basis function (RBF)
Penalty, C	10
Gamma, γ	0.3
Classification scheme	One-vs-one

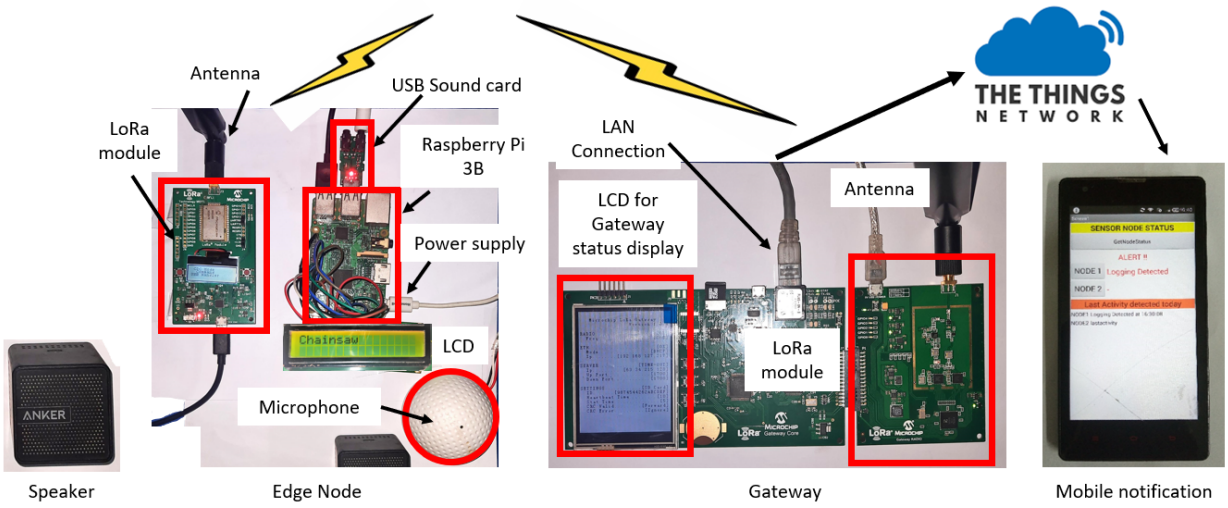


Fig. 5. IASN-based laboratory prototype.

a pop-up notification. The developed laboratory prototype for the audio signal-based smart forest monitoring application is depicted in Fig. 5.

B. Implementation of the Proposed Efficient Method

The SVM classifier is trained on the proposed features extracted from the training data listed in Table I. For the proposed feature extraction from the audio signals, the audio signal is first decomposed using STHT to obtain a spectrogram. The input audio signal is segmented into small frames of length 512 sample points, that is, 32 ms of the audio signal when sampled at a 16-kHz sampling rate with 50%, that is, 256 sample points or 16 ms overlapping between adjacent frames. The resulting number of frames for a 1-s audio clip is $N = 63$. Each frame is transformed into a frequency domain using 512 points FHT, resulting in a 512×63 spectrogram. The proposed LLFs are extracted from this spectrogram. These extracted features are provided as input to train the RBF kernel-based SVM classifier. During the SVM's training phase, the value of gamma (γ) is searched between [0.1 1] and that of penalty (C), that is, the cost function between [1 100]. The one-versus-one training scheme of the SVM classifiers is incorporated. Once the SVM classifier model is trained, it is loaded into the storage of the Raspberry Pi for laboratory experiments purposes. The hyperparameters for implementing the proposed methodology using the developed IASN system are listed in Table V.

The laboratory experiment is conducted on the test dataset listed in Table I. The total duration of each of the categories is 1) chainsaw, ~10 h; 2) handsaw, 1 h 52 min; 3) vehicle, 5 h 8 min; 4) speech, 5 h 50 min; and 5) forest ambience, 8 h 41 min. These pre-recorded audio signals are regenerated using a loudspeaker inside the laboratory environment to create pseudo-real-time illegal logging activities sound and forest ambience. The audio acquisition unit of the developed IASN prototype receives this sound and converts it into a digital signal. The noise gating unit improves the SNR before extracting the proposed LLFs. These extracted features are provided as input to the trained SVM classifier, which classifies it into one of the five considered classes. Upon detection of a threatening class, the sensor node generates an alert message, which is sent to the LoRa gateway using the LoRa communication module. The LoRa-to-IP gateway re-transmits the received message to the Things network cloud application using an ethernet connection. The mobile device fetches this message from the cloud application as a real-time pop-up notification. This facilitates the real-time remote monitoring and detection of illegal logging activities in forests. The predicted output by the embedded platform is saved in a text file to obtain a confusion matrix, as shown in Table VI. The classification accuracy is estimated using the average accuracy obtained for each class.

TABLE VI
CONFUSION MATRIX OF DEVELOPED IASN SYSTEM

Correct classification instance: 112813						Accuracy: 96.61%
Misclassified instance: 3883						Error rate: 3.39%
	Ground Truth					
Prediction	C	H	V	S	F	∑
C	38408	165	63	99	250	38985
H	302	6346	8	51	78	6785
V	169	35	17759	69	556	18580
S	35	124	98	20544	629	21430
F	286	85	552	237	29756	30916
∑	39192	6755	18480	21000	31269	116696

C: Chainsaw, H: Handsaw, V: Vehicle, S: Speech, F: Forest ambience

C. Performance Metrics

A classifier's performance is measured by the number of true positives (T_P), true negatives (T_N), false positives (F_P), and false negatives (F_N). Following the estimation of the confusion matrix, five metrics have been produced to measure the classifier's classification performance: accuracy ($A_C = (T_P + T_N)/(T_P + T_N + F_P + F_N)$) as an overall measure, error rate ($E_R = 100 - A_C$), sensitivity ($S_E = (T_P/(T_P + F_N))$) refers to the proportion of logging events identified as meeting the logging event, specificity ($S_P = (T_N/(T_N + F_P))$) is the proportion of correctly identified

TABLE VII
PERFORMANCE METRIC

Class	TP	FP	FN	A_C	S_E	S_P	F_S
C	38408	577	784	98.83	97.99	99.25	98.25
H	6346	439	409	99.27	93.94	99.60	93.73
V	17759	821	721	98.67	96.09	99.16	95.83
S	20544	886	456	98.85	97.82	99.07	96.83
F	29756	1160	1513	97.71	95.16	98.64	95.70
Total	**112813**	**3883**	**3883**	**98.67**	**96.20**	**99.14**	**96.07**

The value of A_C, S_E, S_P and F_S are in per cent.

TABLE VIII
COMPARISON WITH REPORTED WORK

Study [Ref.]	Classes	Method	Accuracy (%)
Ahmad et al. [22]	2 (Chainsaw, Background)	LLF + Distance	92%
Sharma [25]	2 (Chainsaw, Background)	MMFCC + SBGC	90.52%
Czuni et al. [17]	2 (Chainsaw, Background)	TESPAR + Distance	79.5%
Kalhara et al. [21]	2 (Chainsaw, Background)	Spectrogram + CNN	96%
Andreadis et al. [27]	4 (Chainsaw, Handsaw, Fire, Background)	MFCC + SVM	85.37%
Mporas et al. [23]	2 (Chainsaw, Background)	LLF + MFCC + SVM	94.42%
Proposed	**5 (Chainsaw, Handsaw Vehicle, Speech, Forest ambience)**	**STHT+LLF(16)+SVM**	**96.61%**

LLF: Low-Level Features, MMFCC: Modified Mel Frequency Cepstral Coefficients, SBGC: Spectral feature based Gauss-Bayesian Classifier, TESPAR: Time Encoded Signal Processing And Recognition, CNN: Convolutional Neural Network, SVM: Support Vector Machines, STHT: Short Time Hartley Transform.

non-logging or forest ambience events, and F-score ($F_S = (2T_P/(2T_P + F_N + F_P))$).

V. RESULTS AND DISCUSSION

The laboratory experiments are conducted in real time using the proposed methodology, and the results of the developed system are represented in the confusion matrix in Table VI. In contrast, the column of the confusion matrix denotes the ground truth of the classified instances. In a total of 116 696 testing audio clips, 112 813 clips are correctly identified by the developed IASN system, reporting a high accuracy of 96.61% along with an error rate of only 3.39%. However, in Table VI, it is observed that the accuracy of the handsaw ("H") class, that is, 93.94%, is less compared to the other audio class, which is due to the fewer data of this class available to train the classifier. A high overall value of S_E, S_P, and F_S parameters is obtained, which are reported to be 96.20%, 99.14%, and 96.07%, respectively, as presented in Table VII. Further, considering the real-world forest settings, if the audio clip of any class apart from the classes considered in this study is recorded by the audio acquisition unit, then this audio signal is classified into the forest ambience category.

The hardware implementation of the proposed methodology has also been analyzed for energy consumption and timing. The energy consumption is recorded every ten minutes for 100 min, and then the average energy consumption is computed, which is reported to be as low as 0.184 mWh. For the timing analysis, 4162 one-second audio recordings from the testing dataset had been processed and classified in 100 min. In addition, the execution time to process and classify a 1-s audio signal is 441 ms. The energy and timing analysis, along with other specifications of the developed IASN system, is summarized in Table IX.

A brief comparison is presented in Table VIII between the results achieved by the proposed methodology on the IASN system and the existing work available in the literature. A fair comparison is quite tedious to make, which is due to the fact that different works have been evaluated on distinct databases with uneven numbers and classes of the audio signal for illegal logging detection. Table VIII concludes that the proposed method with the developed hardware platform has reported the best performance among the available methodologies reported in the literature. The number of audio signals, that is, clips, varies greatly among the various categories of the audio signal in training and testing datasets. The accuracy reported for a particular class can be considered more reliable and significant. As the work presented in this article achieves a higher accuracy than the existing works, it is implicit in concluding that the features extracted in the time–frequency domain using the proposed STHT-based LLF extraction technique are efficient and significant in discriminating between various categories of an audio signal when combined with the SVM classifier. Further, Table IX presents a comparison based on features of the proposed IASN system in contrast to the reported systems in the literature. In [10], a monitoring system consists of audio and vibration sensors, a data acquisition unit, ATmega328P, and Raspberry Pi for processing audio signals and a ZigBee wireless module for data transmission. The detection is purely threshold-based, with a resolution of 10 s. The accuracy reported is only 90% for the binary classification problem, and no alert system is employed. In [12], a system consisting of Arduino UNO as a computing platform and microphone and accelerometer sensors are reported. The illegal activity detection is based on threshold comparison. The alert message and location information are sent to the server using LoRa communication and GPS module. The end user receives real-time notifications on a mobile application, that is, Telegram. In [21], Arduino UNO and Raspberry Pi are used as computing platforms. The nRF24L01 and Wi-Fi are used for wireless data transmission. The sensor used is a microphone. A spectrogram with the CNN classifier method is employed for detecting illegal logging. The system reports an accuracy of 98% but for binary class classification and 3-s resolution. Also, it does not employ an alert notification system. In [27], an ARM Cortex M4F-based system is developed with a microphone as a sensor and a LoRa module for wireless communication. An MFCC feature and the CNN classifier are implemented for four classes for illegal logging detection. However, the accuracy reported is only 85% with 4-s resolution, and no alert notification system is employed. The hardware developed in this article addresses these limitations. From Table IX, it can be concluded that the proposed system is more accurate and capable of detecting a greater number of illegal logging classes in comparison to the existing works. Further, an alert notification is received on the mobile upon the detection of any illegal logging activity. Therefore, the proposed system has more features and offers a better service in remote forest monitoring applications. The proposed method implementation using an edge computing-based embedded platform provides an enriched interface to the forest authorities regarding real-time remote forest monitoring. It maintains a

TABLE IX
COMPARISON WITH REPORTED HARDWARE SYSTEMS

Features	Device A Ref. [10]	Device B Ref. [12]	Device C Ref. [21]	Device D Ref. [27]	Proposed
Analysis	Real-time	Real-time	Real-time	Real-time	**Real-time**
Platform	ATmega32P Raspberry Pi	Arduino UNO	Arduino UNO Raspberry Pi	ARM Cortex M4F	**Raspberry Pi**
Wireless	ZigBee	LoRa GPS	nRF24L01 WiFi	LoRa	**LoRa**
Sensor	Microphone Accelerometer	Microphone Accelerometer	Microphone	Microphone	**Microphone**
Method	Thresholding	Thresholding	Spectrogram CNN	MFCC + CNN	**STHT + LLF + SVM**
Resolution	10 s	-	3 s	4 s	**1 s**
Accuracy	90%	-	96%	85%	**96.61%**
Classes	3	3	2	4	**5**
Processing Time	-	-	-	-	**441 ms**
Energy Consumption (per second)	-	-	-	-	**0.184 mWh**
Alert system	No	Web Application Telegram notification	No	No	**Mobile notification**

database in the cloud server for later analysis and improvement in an existing monitoring system. The proposed approach allows remote real-time monitoring and detection of illegal logging activities in the forest to provide preventive measures at the earliest. This implementation will provide assistive forest monitoring to the forest guards/rangers, thereby improving the protection of endangered forests.

VI. CONCLUSION

This article proposes an STHT-based low-level features extraction from the pre-recorded data from acoustic sensors and its classification using an SVM classifier to monitor and detect illegal logging activities. The low-level features are extracted from the spectrogram of an acoustic sensor data obtained by decomposition in the time–frequency domain having low computational complexity. The overall accuracy reported is 96.61%, along with sensitivity, specificity, and F-score of 96.20%, 99.14%, and 96.07%, respectively. Further, the proposed methodology is prototyped on a developed COTS-based microprocessor platform to facilitate real-time monitoring and generation of alerts. The proposed platform is validated in the laboratory experimental setup using the regenerated sounds from acoustic sensors of the considered pre-recorded audio classes on a loudspeaker. Upon detection of any logging activity, IASN sends the alert message to the mobile device as a pop-up notification in real time. Therefore, the developed IASN system can be deployed in forests for real-time remote monitoring and generation of quick alerts upon detection of any illegal logging activity.

REFERENCES

[1] *Survey Report*. Accessed: Jul. 29, 2022. [Online]. Available: https://www.fao.org/forest-resources-assessment/2020/en/

[2] *Survey Data*. Accessed: Jul. 29, 2022. [Online]. Available: https://data.worldbank.org/indicator/

[3] K. P. Chethan, J. Srinivasan, K. Kriti, and K. Sivaji, "Sustainable forest management techniques," in *Deforestation Around the World*, P. Moutinho, Ed., London, U.K.: IntechOpen, 2012.

[4] C. A. Okia, Ed., "Deforestation: Causes, effects and control strategies," in *Global Perspectives on Sustainable Forest Management*. Rijeka, Croatia: InTech, 2012.

[5] *Survey Data*. Accessed: Jul. 29, 2022. [Online]. Available: https://www.globalforestwatch.org/dashboards/global/

[6] M. Babis, M. Duricek, V. Harvanova, and M. Vojtko, "Forest guardian—Monitoring system for detecting logging activities based on sound recognition, researching solutions in artificial intelligence, computer graphics and multimedia," in *Proc. IIT SRC*, Bratislava, Slovakia, May 2011, pp. 1–6.

[7] W. B. Magrath, R. L. Grandalski, G. L. Stuckey, G. B. Vikanes, and G. R. Wilkinson, "Timber theft prevention: Introduction to security for forest managers," Sustain. Development-East Asia Pacific Region, World Bank Publication, Washington, DC, USA, 2007.

[8] M. J. Baucas and P. Spachos, "Using cloud and fog computing for large scale IoT-based urban sound classification," *Simul. Model. Pract. Theory*, vol. 101, May 2020, Art. no. 102013.

[9] K. T. Chui, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, "An accurate ECG-based transportation safety drowsiness detection scheme," *IEEE Trans. Ind. Informat.*, vol. 12, no. 4, pp. 1438–1452, Aug. 2016.

[10] J. T. Chen, C. B. Lin, J. J. Liaw, and Y. Y. Chen, "Improving the implementation of sensor nodes for illegal logging detection," in *Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Sendai, Japan. Cham, Switzerland: Springer, 2018, pp. 212–219.

[11] G. A. Mutiara, N. Suryana, and O. Mohd, "Multiple sensor on clustering wireless sensor network to tackle illegal cutting," *Int. J. Adv. Sci., Eng. Inf. Technol.*, vol. 10, no. 1, pp. 164–170, Feb. 2020.

[12] Y.-Y. Chen and J.-J. Liaw, "A novel real-time monitoring system for illegal logging events based on vibration and audio," in *Proc. IEEE 8th Int. Conf. Awareness Sci. Technol. (iCAST)*, Taiwan, Nov. 2017, pp. 470–474.

[13] G. A. Mutiara, N. S. Herman, and O. Mohd, "Using long-range wireless sensor network to track the illegal cutting log," *Appl. Sci.*, vol. 10, no. 19, p. 6992, Oct. 2020.

[14] D. C. Prasetyo, G. A. Mutiara, and R. Handayani, "Chainsaw sound and vibration detector system for illegal logging," in *Proc. Int. Conf. Control, Electron., Renew. Energy Commun. (ICCEREC)*, Bali, Indonesia, Dec. 2018, pp. 93–98.

[15] V. Yaremenko, M. R. Azimi-Sadjadi, and J. Zacher, "Unattended acoustic sensor systems for source detection, classification, and tracking," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 344–354, Feb. 2019.

[16] F. Pianegiani, M. Hu, A. Boni, and D. Petri, "Energy-efficient signal classification in ad hoc wireless sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 1, pp. 190–196, Jan. 2008.

[17] L. Czúni and P. Z. Varga, "Time domain audio features for chainsaw noise detection using WSNs," *IEEE Sensors J.*, vol. 17, no. 9, pp. 2917–2924, May 2017, doi: 10.1109/JSEN.2017.2670232.

[18] M. V. Ghiurcau, C. Rusu, R. C. Bilcu, and J. Astola, "Audio based solutions for detecting intruders in wild areas," *Signal Process.*, vol. 92, no. 3, pp. 829–840, Mar. 2012.

[19] Z. Liu, J. Huang, Y. Wang, and T. Chen, "Audio feature extraction and analysis for scene classification," in *Proc. 1st Signal Process. Soc. Workshop Multimedia Signal Process.*, Jun. 1997, pp. 343–348, doi: 10.1109/MMSP.1997.602659.

[20] A. Gaita, G. Nicolae, A. Radoi, and C. Burileanu, "Chainsaw sound detection based on spectral Haar coeffluents," in *Proc. Int. Symp. ELMAR*, Sep. 2018, pp. 139–142.

[21] P. G. Kalhara, V. D. Jayasinghearachchi, A. H. A. T. Dias, V. C. Ratnayake, C. Jayawardena, and N. Kuruwitaarachchi, "TreeSpirit: Illegal logging detection and alerting system using audio identification over an IoT network," in *Proc. 11th Int. Conf. Softw., Knowl., Inf. Manage. Appl. (SKIMA)*, Malabe, Sri Lanka, Dec. 2017, pp. 1–7.

[22] S. F. Ahmad and D. K. Singh, "Automatic detection of tree cutting in forests using acoustic properties," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 3, pp. 757–763, Mar. 2022.

[23] I. Mporas, I. Perikos, V. Kelefouras, and M. Paraskevas, "Illegal logging detection based on acoustic surveillance of forest," *Appl. Sci.*, vol. 10, no. 20, p. 7379, Oct. 2020, doi: 10.3390/app10207379.

[24] S. K. Yadav, K. Tyagi, B. Shah, and P. K. Kalra, "Audio signature-based condition monitoring of internal combustion engine using FFT and correlation approach," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 4, pp. 1217–1226, Apr. 2011.

[25] G. Sharma, "Acoustic signal classification for deforestation monitoring: Tree cutting problem," *J. Comput. Sci. Syst. Biol.*, vol. 11, no. 2, pp. 178–184, 2018.

[26] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Appl. Acoust.*, vol. 158, Jan. 2020, Art. no. 107020, doi: 10.1016/j.apacoust.2019.107020.

[27] A. Andreadis, G. Giambene, and R. Zambon, "Monitoring illegal tree cutting through ultra-low-power smart IoT devices," *Sensors*, vol. 21, no. 22, p. 7593, Nov. 2021.

[28] J. Svatos and J. Holub, "Impulse acoustic event detection, classification, and localization system," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023, doi: 10.1109/TIM.2023.3252631.

[29] S. Raj and K. C. Ray, "ECG signal analysis using DCT-based DOST and PSO optimized SVM," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 3, pp. 470–478, Mar. 2017, doi: 10.1109/TIM.2016.2642758.

[30] V. Singh, K. C. Ray, and S. Tripathy, "Blind detection and classification algorithm for smart audio monitoring system," in *Proc. IEEE Int. Symp. Smart Electron. Syst. (iSES)*, Dec. 2019, pp. 133–138.

[31] R. V. L. Hartley, "A more symmetrical Fourier analysis applied to transmission problems," *Proc. IRE*, vol. 30, no. 3, pp. 144–150, Mar. 1942.

[32] R. N. Bracewell, "Discrete Hartley transform," *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1832–1835, 1983.

[33] J.-C. Liu and T. P. Lin, "Short-time Hartley transform," *IEE Proc. F, Radar Signal Process.*, vol. 140, no. 3, p. 171, 1993.

[34] R. N. Bracewell, "The fast Hartley transform," *Proc. IEEE*, vol. 72, no. 8, pp. 1010–1018, Aug. 1984, doi: 10.1109/PROC.1984.12968.

[35] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2000.

[36] Accessed: Jul. 2, 2022. [Online]. Available: https://freesound.org/search/?q=handsaw

[37] Accessed: Jul. 2, 2022. [Online]. Available: https://freesound.org/search/?q=chainsaw

[38] Accessed: Jul. 2, 2022. [Online]. Available: https://freesound.org/search/?q=engine

[39] Accessed: Jul. 2, 2022. [Online]. Available: https://freesound.org/search/?q=speech

[40] Accessed: Jul. 2, 2022. [Online]. Available: http://research.google.com/audioset/

[41] B. Wang and W. Wong, "Real time hearing enhancement in crowded social environments with noise gating," *Speech Commun.*, vol. 99, pp. 173–182, May 2018, doi: 10.1016/j.specom.2018.03.010.

[42] W. A. Sethares, R. D. Morris, and J. C. Sethares, "Beat tracking of musical performances using low-level audio features," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 2, pp. 275–285, Mar. 2005.

[43] D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, and L.-H. Cai, "Music type classification by spectral contrast feature," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 1, Aug. 2002, pp. 113–116, doi: 10.1109/ICME.2002.1035731.

[44] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *J. Acoust. Soc. Amer.*, vol. 130, no. 5, pp. 2902–2916, Nov. 2011.

[45] N. Nehrebecka, "Predicting the default risk of companies. Comparison of credit scoring models: Logit vs support vector machines," *Econometrics*, vol. 22, no. 2, pp. 54–73, Jun. 2018.

[46] M. Magno, F. Vultier, B. Szebedy, H. Yamahachi, R. H. R. Hahnloser, and L. Benini, "A Bluetooth-low-energy sensor node for acoustic monitoring of small birds," *IEEE Sensors J.*, vol. 20, no. 1, pp. 425–433, Jan. 2020, doi: 10.1109/JSEN.2019.2940282.

[47] J. Lopez-Ballester, A. Pastor-Aparicio, S. Felici-Castell, J. Segura-Garcia, and M. Cobos, "Enabling real-time computation of psycho-acoustic parameters in acoustic sensors using convolutional neural networks," *IEEE Sensors J.*, vol. 20, no. 19, pp. 11429–11438, Oct. 2020, doi: 10.1109/JSEN.2020.2995779.

[48] M. M. Faraji, S. B. Shouraki, E. Iranmehr, and B. Linares-Barranco, "Sound source localization in wide-range outdoor environment using distributed sensor network," *IEEE Sensors J.*, vol. 20, no. 4, pp. 2234–2246, Feb. 2020.

[49] J. Cen et al., "A mask self-supervised learning-based transformer for bearing fault diagnosis with limited labeled samples," *IEEE Sensors J.*, vol. 23, no. 10, pp. 10359–10369, May 2023, doi: 10.1109/JSEN.2023.3264853.