Energy-Efficient IoT-Fog-Cloud Architectural Paradigm for Real-Time Wildfire Prediction and Forecasting

Harkiran Kaur and Sandeep Kumar Sood

Abstract—**Wildfires are catastrophic disasters. They pose a fatal threat not only to the forest resources but also to the entire regime of flora and fauna, gravely disturbing the bio-diversity and ecology of the region. The frequency and severity of wildfires are expected to grow, owing to global warming. Therefore, it is essential to adopt a comprehensive, multifaceted approach that enables the real-time monitoring of forest terrains and prompt responsiveness. The Internet of Things (IoT) technology has grown exponentially in recent years, with IoT sensors being deployed to monitor and collect time critical data. This research proposes an integrated IoT-Fog-Cloud energy-efficient framework for wildfire prediction and forecasting. Initially, analysis of variance and Tukey's *post hoc* test-based energy conserving mechanism ensures the enhanced lifetime of resource-constrained sensors by adapting the sampling rate of wildfire influent parameters (WIPs) at fog layer. Principal component analysis (PCA) is employed for WIPs' reduction. Wildfire vulnerability level of a forest terrain is predicted and forecasted using Naïve Bayes (NB) classifier and seasonal auto regressive integrated moving average model, respectively, at cloud layer. Burnt forest area is also predicted using support vector regression. The implementation results of the proposed framework prove its efficiency in predicting and forecasting wildfires.**

Index Terms—**Analysis of variance (ANOVA), fog computing, Internet of Things (IoT), seasonal auto-regression integrated moving average (SARIMA), support vector regression (SVR), Tukey's *post hoc* test.**

I. Introduction

FORESTS are an indispensable asset for humanity. They are the protectors of earth's ecological balance. Wildfires are a great menace to the forests across the globe. Large swathes of forests are devoured by wildfires every year, causing damages beyond measure and description to the human race and natural assets. Deviations in local weather patterns, global warming, and extinction of rare species of flora and fauna are amongst the long-term disastrous consequences of wildfires. In 2017, a total of 71 499 wildfires burnt a massive forest area of 10 million acres [1]. Approximately 8000 wildfires occur each year in Canada, gutting an average forest area of about 2.5 million ha/year. The threat of wildfires is expected to increase globally in both frequency and intensity, fuelled by global warming and drought conditions. Therefore, real-time monitoring of wildfire-prone forest terrains for timely prediction and detection of wildfires can significantly dilute the destruction caused by them.

Several technologies and systems have been used over the years for the surveillance of wildfire-prone terrains. These include fire watch towers, satellite monitoring, and optical cameras centric systems. Unreliability of fire watch towers, accompanied by the challenging working and living conditions of fire lookout personnel, led to the adoption of satellite systems and imagery. However, these systems are not suitable for real-time monitoring due to lengthy scanning cycle of 2 days, besides suffering from resolution constraints. Optical cameras centric systems are profoundly affected by topography and weather conditions such as direct and intense sunlight and clouds. Also, these systems are prone to false alerts.

As a promising substitute, IoT is an emerging technology that can be adopted for real-time wildfire monitoring. IoT-enabled wildfire surveillance involves the deployment of small, cost-effective, battery-powered IoT sensors for assessing time-critical wildfire influent parameters (WIPs). The data captured by the sensors need to be analyzed with the minimum delay because wildfires ignite, spread, and assume an alarming proportion in short period. Real-time analysis of WIPs is accomplished by the use of fog computing. Fog computing brings cloud computing paradigm to the edge of the network and offers several advantages such as low latency, location awareness, quality of service assurance, and real-time analysis and alert generation [2], [3]. Therefore, fog computing, in association with cloud computing, aids in the early prediction and swift detection of wildfires along with real-time alert generation to the forest department and disaster response agencies. The layered architecture of fog computing is depicted in Fig. 1. The proposed system integrates IoT, fog computing, cloud computing, along with data mining and analysis techniques for real-time wildfire prediction and forecasting. The objectives of the proposed system comprise the following:

1) IoT-enabled surveillance of wildfire-prone forest terrains for WIPs;
2) enhancement of sensor network lifespan through efficient energy utilization of battery-constrained sensor nodes;

Manuscript received February 12, 2019; revised June 12, 2019; accepted June 15, 2019. Date of publication July 2, 2019; date of current version June 3, 2020. *(Corresponding author: Harkiran Kaur.)*

The authors are with the Guru Nanak Dev University Regional Campus Gurdaspur, Gurdaspur 143521, India (e-mail: harkiran1992@gmail.com; san1198@gmail.com).

Digital Object Identifier 10.1109/JSYST.2019.2923635

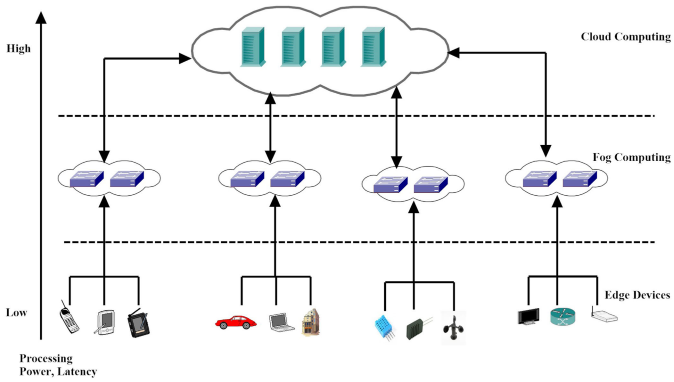


Fig. 1. Layered architecture fog computing.

3) wildfire vulnerability prediction and forecasting of forest terrains;
4) burnt area prediction in the event of wildfire occurrence;
5) sharing of the analysis results with forest department and disaster-response agencies.

This paper is organized into following sections. Section II reviews some of the relevant literature in wildfire monitoring domain. The proposed system for wildfire surveillance, prediction, and forecasting is demonstrated in Section III. Implementation analysis of the proposed system is presented in Section IV. Finally, Section V concludes this paper.

II. Related Work

Integration of IoT, fog computing, cloud computing, and data mining and decision-making techniques for real-time wildfire surveillance, prediction, and forecasting has not been delved yet to the best of our knowledge. Different proposals have been put forward for real-time wildfire surveillance and detection based on wireless sensor networks (WSNs) over the years. This section gives insights into some of the highly relevant studies.

In 2012, Aslan *et al.* [4] proposed a comprehensive framework for wildfire surveillance and detection using WSNs. The framework aims to detect a wildfire risk as early as possible while considering the efficient energy consumption of the battery-constrained sensor nodes. This study examined only two parameters, i.e., temperature and relative humidity for wildfire detection using threshold-based decision making, in contrast to 11 parameters considered in our proposed framework. Moreover, wildfire forecasting and burnt area prediction facets of wildfire surveillance are not explored in this paper. In 2014, Ulucinar *et al.* [5] developed a WSN data harvesting application for wildfire detection, comprising of sensor nodes equipped with temperature and relative humidity sensors organized into clusters. Temperature variation based fire detection algorithm is also proposed. The proposal only considers one parameter, i.e., temperature, for wildfire detection. Also, energy consumption and forecast capability issues are not discussed.

In 2016, Molina-Pico *et al.* [6] discussed a WSN comprising of a combination of central (CNs) and standard sensor nodes (SNs). Diverse parameters such as temperature, humidity, CO, CO_2, precipitation, and wind's speed and direction are captured by SNs. The CNs are characterized by powerful communication and computation abilities in comparison to the sensor nodes. Their main task is to accumulate and cluster data from the sensor nodes, manage alerts and commands, and build the core network. On the power autonomy front, only the central nodes are equipped with solar power harvesting facility. In 2017, Garcia-Jimenez *et al.* [7] proposed a fuzzy system approach based on overlap indices for forest fire detection. However, the study solely emphasizes on wildfire detection, without focusing on an essential aspect of efficient energy consumption of the resource-constrained sensor nodes. In 2018, Lin *et al.* [8] proposed a fuzzy inference and big data analysis algorithm for wildfire prediction. The use of wireless rechargeable sensor network with solar energy harvesting is the most significant contribution of this paper.

In 2018, Toledo-Castro *et al.* [9] put forward a proposal for real-time environmental monitoring of dynamic wildfire risk factors by using WSNs and novel IoT technologies. Weather-related variables, polluting gases, and oxygen level are captured in real time to estimate the existence of wildfire risks in the short-term. Moreover, they employed analytic hierarchy process decision making method to determine the level of fire spread, and when essential, environmental alerts are sent. The protection of integrity, confidentiality, and authenticity of ecological information and alerts with implementations of Lamports authentication scheme, Diffie–Lamport signature, and AES-CBC block cipher is a crucial aspect of this study.

All the studies discussed above focus on either one or only few aspects of wildfire surveillance. This paper incorporates all the key aspects of wildfire surveillance in a single framework, i.e., wildfire detection, prediction, forecasting, and burnt area prediction, while focusing on the efficient utilization of the energy of resource-constrained sensor nodes.

III. Proposed System

The three-tier architecture of the IoT-empowered fog-aided energy-conserving system for smart wildfire surveillance, prediction, and forecasting is depicted in Fig. 2. The proposed system encompasses three layers, namely, data sensing layer, fog layer, and cloud layer. Each layer accomplishes its essential task, thereby providing necessary services to the adjacent layers. The data sensing layer consists of sensor nodes responsible for the surveillance of the fire-prone forest terrains for WIPs. The recorded values of the WIPs are, then, relayed to the fog layer. The WIPs are analyzed at fog e-nodes for data variance to adapt the sampling rate of the sensor nodes. This leads to the efficient energy consumption of resource-constrained sensor nodes.

Furthermore, the fog layer is responsible for reducing the dimensionality of WIPs under consideration, before transmitting them to the cloud layer through gateways. The cloud layer determines the present wildfire vulnerability of a forest terrain, as well as estimates future wildfire vulnerability trends. Finally, the compiled WIPs data, along with prediction and forecasting results, are stored in the cloud layer, thus aiding the forest department and disaster management agencies in

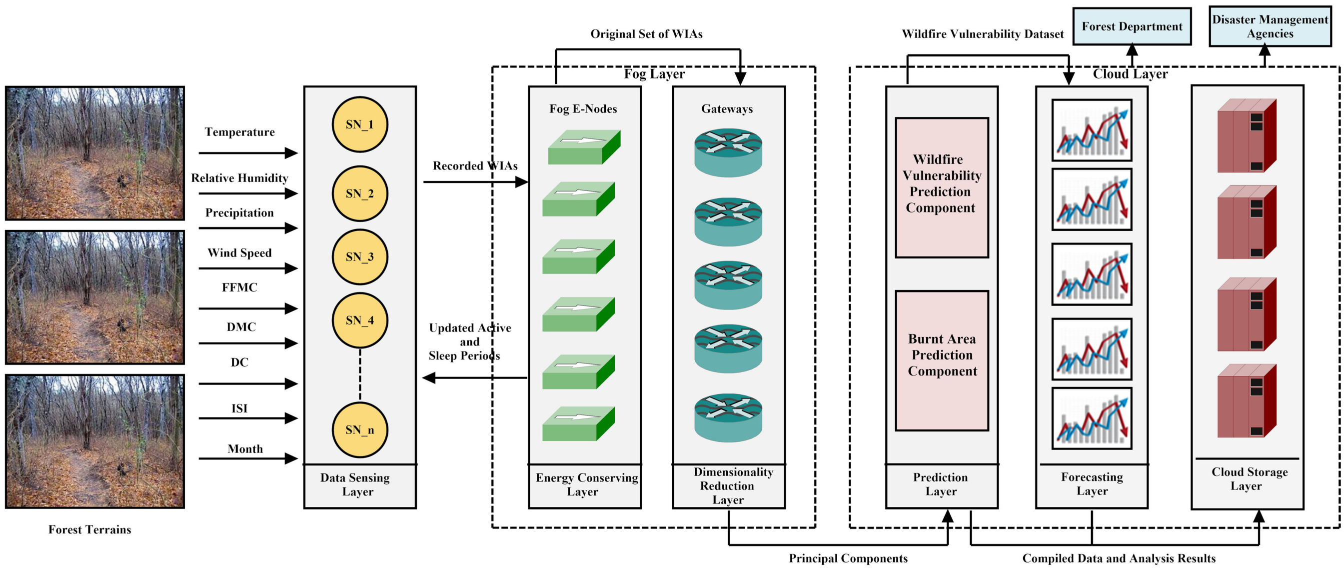


Fig. 2. Layered architecture of the proposed system.

TABLE I
WILDFIRE INFLUENT PARAMETERS

S.No.	Wildfire Influent Parameters (WIPs)	Description
1.	Temperature	Air temperature of the forest teraain being monitored.
2.	Relative Humidity	Amount of water vapour present in air of the forest terrain being monitored.
3.	Precipitation	Amount of rainfall, snowfall, hail and other forms of precipitation.
4.	Wind Speed	Rate of the movement of wind in distance per unit of time.
5.	Month	Month of the year.
6.	FFMC (Fine Fuel Moisture Code)	Numeric rating of the moisture content of litter and other cured fine fuels. This code is an indicator of relative ease of ignition and the flammability of fine fuel.
7.	DMC (Duff Moisture Code)	Numeric rating of the average moisture content of loosely compacted organic layers of moderate depth.
8.	DC (Drought Code)	Numeric rating of the average moisture content of deep, compact organic layers. This code is an indicator of seasonal drought effects on forest fuels.
9.	ISI (Initial Spread Index)	Numeric rating of the expected rate of fire spread. It combines the effects of wind and the FFMC on rate of spread.

effective mitigation and management of wildfires. A detailed description of the functionalities and services of each layer is given ahead.

A. Data Sensing Layer

Data sensing layer performs the task of monitoring the forest terrain for assessing the time-critical WIPs. Data are retrieved ubiquitously from different sensor nodes embedded in the forest terrain. These sensor nodes work on wireless sensing principle and are competent in sensing and transmitting data in real time. Various meteorological sensors such as temperature sensor, precipitation sensor, humidity sensor, and wind speed sensor, are integrated into each sensor node. Also, location attributes specifying the spatial coordinates of the potential epicenter of wildfire also play a crucial role in mitigating the disastrous effects of wildfires. Therefore, the data accumulated by sensor nodes are categorized into two datasets, which are as follows.

1) WIPs dataset: This dataset encompasses various meteorological attributes, namely temperature, relative humidity, precipitation, and wind speed. Besides, month, fine fuel moisture code (FFMC), duff moisture code (DMC), drought code (DC), and initial spread index (ISI), which play a crucial role in wildfire outbreak and its consequent spread, form an integral part of this dataset.
2) Location dataset: This dataset consists of spatial coordinates of a probable site of the fire, i.e., the x-axis and y-axis spatial coordinates. Different WIPs, along with their description, are illustrated in Table I. The acquired data are relayed to the fog layer for real-time analysis.

B. Fog Layer

The fog layer acts as a bridge between the data sensing layer and the cloud layer. It is responsible for the real-time analysis and processing of the data transmitted to it by the data sensing layer. In the proposed system, the fog layer comprises of two sub-layers, namely energy conserving layer and dimensionality reduction layer.

1) Energy Conserving Layer: Sensors-enabled wildfire surveillance demands continuous monitoring of WIPs prevalent in forest terrain. Therefore, periodic sensors are well suited for sensing and transmitting WIPs such as temperature, precipitation, relative humidity, and wind speed periodically. However, periodic sensors face two significant challenges. First, the sensors must have a lifespan long enough to fulfill the wildfire surveillance requirements [10]. Second, massive data concerning WIPs accumulated by sensors make data management a complex task. Thus, the energy conserving layer is entrusted with the responsibility of minimizing the amount of data sensed and transmitted by the sensors without significant loss in accuracy. This would promote the optimal energy consumption of battery-constrained sensors, thereby leading to their enhanced lifespan.

The analysis of variance (ANOVA) model, in association with Tukey's *post hoc* test, is used for adapting the sampling rate of sensor nodes in accordance with the changing dynamics of WIPs. In the case of periodic sensors, each time period t_p includes of several time slots t_s. Every sensor node SN_k reports a new WIPs' reading w_k in each time slot t_s of the time period t_p, thus resulting in a vector of readings denoted as

$$W_k = \{w_1, w_2, w_3, \ldots, w_{T-1}, w_T\} \tag{1}$$

where T represents the total number of WIPs readings logged in the time period tp. As previously specified, WIPs fluctuate slowly. The one-way ANOVA model is used to assess the gross variation (GV) of the WIPs recorded by a sensor node SN_k in Ntime periods. The $GV$is defined as the sum of the variation within a period or intraperiod variation (V_{intra}) and the variation between periods or interperiod variation (V_{inter}) denoted as

$$GV = V_{\text{intra}} + V_{\text{inter}} \tag{2}$$

$$\sum_{l=1}^{N}\sum_{k=1}^{n_l}(w_{kl}-\mu)^2 = \sum_{l=1}^{N}\sum_{k=1}^{n_l}(w_{kl}-\mu^l)^2 + \sum_{l=1}^{N} n_l \times (\mu^l - \mu)^2 \tag{3}$$

where w_{kl} represents the kth reading captured by the sensor node SN_k in the lth time period; n_l represents the number of readings registered in the lth time period; N represents the total number of periods; n_N represents the total number of readings captured in N periods; μ^l represents the average of readings recorded in the lth time period; μ represents the average of readings captured in all N time periods.

The outcomes of the one-way ANOVA method are used to determine whether the means of different datasets accumulated in consecutive time periods differ significantly. Also, we extend the one-way ANOVA model to the Tukey's *post hoc* test, which is a widely recognized statistical test to determine whether the variation between the WIPs captured in different periods surpasses a predefined threshold value T_{th}.This is followed by adapting the sampling rate of sensor nodes by updating their active and sleep time periods to facilitate efficient usage of their energy. One has the following:

$$\delta_{\text{total}} = \sum_{k=1}^{n_1} w_{1k}^2 + \sum_{k=1}^{n_2} w_{2k}^2 + \sum_{k=1}^{n_3} w_{3k}^2 + \cdots + \sum_{k=1}^{n_N} w_{Nk}^2 - \frac{\left(\sum_{k=1}^{n_1} w_{1k} + \sum_{k=1}^{n_2} w_{2k} + \sum_{k=1}^{n_3} w_{3k} + \cdots + \sum_{k=1}^{n_N} w_{Nk}\right)^2}{n_N} \tag{4}$$

$$\delta_{\text{among}} = \frac{\left(\sum_{k=1}^{n_1} w_{1\text{k}}\right)^2}{n_1} + \frac{\left(\sum_{k=1}^{n_2} w_{2k}\right)^2}{n_2} + \frac{\left(\sum_{k=1}^{n_3} w_{3k}\right)^2}{n_3} + \cdots + \frac{\left(\sum_{k=1}^{n_N} w_{Nk}\right)^2}{n_N} - \frac{\left(\sum_{k=1}^{n_1} w_{1k} + \sum_{k=1}^{n_2} w_{2k} + \sum_{k=1}^{n_3} w_{3k} + \cdots + \sum_{k=1}^{n_N} w_{Nk}\right)^2}{n_N} \tag{5}$$

$$\delta_{\text{within}} = \delta_{\text{total}} - \delta_{\text{among}} \tag{6}$$

$$\varpi_{\text{among}} = \frac{\delta_{\text{among}}}{N-1} \tag{7}$$

$$\varpi_{\text{within}} = \frac{\delta_{\text{within}}}{n_N - N} \tag{8}$$

$$T = \frac{\varpi_{\text{among}}}{\varpi_{\text{within}}} \tag{9}$$

where δ_{among} denotes the interperiod sum of squares; δ_{within} denotes the intraperiod sum of squares; ϖ_{among} denotes the interperiod mean squares; ϖ_{within} denotes the intraperiod mean squares. Thus, adapting the sampling rate of sensor nodes by updating active and sleep intervals aids in the resourceful consumption of their energy.

The proposed energy conserving layer computes T by using the WIPs recorded by sensor nodes in the two most recent active periods. Based on the calculated value of T, next active and sleep periods of the sensor nodes are determined. If T exceeds T_{th}, the active and sleep periods are set to constant values τ_a and τ_s, respectively, as

$$\tau_{l+1}^{\text{Active}} = \tau_a \tag{10}$$

$$\tau_{l+1}^{\text{Sleep}} = \tau_s. \tag{11}$$

If T is less than T_{th}, the active period is decreased and the sleep period is increased as

$$\tau_{l+1}^{\text{Active}} = \tau_a\left(\frac{T_{\text{th}} - T}{T_{\text{th}}}\right) \tag{12}$$

$$\tau_{l+1}^{\text{Sleep}} = \tau_s\left(\frac{T_{\text{th}} + T}{T_{\text{th}}}\right) \tag{13}$$

where τ_l^{Active} denotes the lth active period and τ_l^{Sleep} the lth sleep period. Also, wakeup signals are sent to the sensor nodes in sleep mode. The working of the energy conserving layer is illustrated in Algorithm 1.

2) Dimensionality Reduction Layer: The enormity of the raw data generated by the sensor nodes deployed for wildfire monitoring makes data analysis and processing a challenging task. Dimensionality reduction techniques are used for transforming

Algorithm 1: Working of Energy Conserving Layer.

Input: Sensor node readings of the two most recent periods *l* and *l-1*, τ_a, τ_s, T_{th} and present_mode
Output: Succeeding active and passive time periods i.e. τ_{l+1}^{Active} and τ_{l+1}^{Sleep}

for *each sensor node* SN_k **do**
  Determine the readings captured during previous two periods *l* and *l-1*
  if *present_mode= "active"* **then**
    Compute T using equations (4) to (9)
    if $T \geq T_{th}$ **then**
      $\tau_{l+1}^{Active} \leftarrow \tau_a$; $\tau_{l+1}^{Passive} \leftarrow \tau_p$
    else
      $\tau_{l+1}^{Active} \leftarrow \tau_a\left(\frac{T_{th}-T}{T_{th}}\right)$; $\tau_{l+1}^{Passive} \leftarrow \tau_s\left(\frac{T_{th}+T}{T_{th}}\right)$
  if *present_mode= "sleep"* **then**
    Send Wake Up Signal after τ_l^{Sleep} time interval.
return τ_{l+1}^{Active} *and* τ_{l+1}^{Sleep}

a dataset that has large dimensions into data with smaller dimensions, while ensuring no or negligible difference in predictive accuracy. In the proposed framework, the principal component analysis (PCA) [11] is employed for reducing WIPs into a smaller set of attributes by removing extraneous, noisy, and redundant WIPs. The PCA transforms the *n* original WIPs, w_1, w_2 ....w_n, into *m* new variables *(m < n)*, $PC_1,\ PC_2$ ..... PC_m, known as principal components (PCs) such that the new variables are uncorrelated with each other. The PCs are represented as

$$PC_i = \alpha_{i1}w_1 + \alpha_{i2}w_2 + \cdots + \alpha_{in}w_n \tag{14}$$

where PC_i denotes the *i*th PC; α_{in} denotes the weight or loading score of the *n*th WIP in PC_i as calculated by the PCA; w_n denotes the recorded value for the *n*th WIP. Each PC is directed toward the maximum variance, excluding the variance already accounted for in all its preceding components. Subsequently, the first component covers the maximum variance, and each component that follows it covers a lesser value of variance. Thus, the first few PCs that project the majority of the information and variability are selected. Consider a WIPs dataset denoted by $W_{l\times n}$, where *l* denotes the total number of instances, and *n* denotes the total number of WIPs under consideration. The PCA transforms this dataset into a new dataset $R_{l\times m}$, where *m* denotes the number of PCs. The PCA constitutes the following five major steps.

(i) Begin with coding the WIPs w_1, w_2,.......w_n to have zero means and unit variance.
(ii) Compute the correlation matrix C.
(iii) Find the eigenvalues $\lambda_1,\ \lambda_2, \ldots,\ \lambda_n$ and the corresponding eigenvectors $v_1,\ v_2, \ldots,\ v_n$ as

$$|C - I\lambda| = 0. \tag{15}$$

(iv) Remove the PCs that only account for a small proportion of the variation in the dataset.
(v) Develop the factor loading matrix and perform a Varimax rotation on the factor loading matrix to infer the PCs.

The reduced data are forwarded to the naïve Bayes' classifier in the prediction layer for further analysis.

C. Cloud Layer

Since the fog layer is not competent enough to store all the data required for long-term monitoring and prediction of wildfire, therefore, the cloud layer is required for its long-term storage, analysis, and processing. It comprises of three sub-layers, namely prediction layer, wildfire forecasting layer, and cloud storage layer, each of which is described ahead.

1) Prediction Layer: The prediction layer comprises two components, namely, wildfire susceptibility prediction component and forest area burnt prediction component.

(a) Wildfire vulnerability prediction component: This component is responsible for assessing the vulnerability of a forest terrain to wildfire outbreak. Naïve Bayes classifier is used for classifying a forest terrain into one of the five vulnerability classes namely, no fire risk (1), low risk (2), medium risk (3), high risk (4), and extremely high risk (5). The probabilistic model of the naïve Bayes classifier is based on Bayes' theorem. Given the values of all the WIPs for a forest terrain, the dimensionality reduction layer computes tuple *R* comprising of the selected PCs. The naïve Bayes classifier is based on the assumption of conditional independence among attributes, which is satisfied as the PCs obtained are uncorrelated to each other. The tuple *R* is considered to be a part of wildfire vulnerability class C_i with the maximum posterior probability given as

$$P\left(\frac{C_i}{R}\right) > P\left(\frac{C_j}{R}\right)\ ; 1 \leq j \leq 5; j \neq i \tag{16}$$

where $P\left(\frac{C_i}{R}\right)$ is the probability that the monitored forest terrain belongs to wildfire vulnerability class C_i when it has measure *R* for the selected PCs and is computed as

$$\mathrm{P}\left(\frac{\mathrm{C_i}}{\mathrm{R}}\right) = \frac{\mathrm{P}\left(\frac{\mathrm{R}}{\mathrm{C_i}}\right)\mathrm{P}\left(\mathrm{C_i}\right)}{\mathrm{P(R)}} \tag{17}$$

where $\mathrm{P}\left(\frac{\mathrm{R}}{\mathrm{C_i}}\right)$ is the probability of having measure *R* for selected PCs when the monitored forest terrain belongs to wildfire vulnerability class C_i and is calculated as

$$\begin{aligned}\mathrm{P}\left(\frac{\mathrm{R}}{\mathrm{C_i}}\right) &= \sum_{\mathrm{k=1}}^{\mathrm{m}} \mathrm{P}\left(\frac{PC_k}{\mathrm{C_i}}\right) \\ &= \mathrm{P}\left(\frac{PC_1}{\mathrm{C_i}}\right) * \mathrm{P}\left(\frac{PC_2}{\mathrm{C_i}}\right) * \ldots.. \mathrm{P}\left(\frac{PC_m}{\mathrm{C_i}}\right)\end{aligned} \tag{18}$$

where *m* is the total number of the selected PCs, and PC_k is the value of the *k*th PC. Algorithm 2 illustrates the working of the naïve Bayes-based wildfire vulnerability prediction component.

(b) Burnt area prediction component: Wildfires lead to burning of massive tracts of forest land. Therefore, predicting the

Algorithm 2: Wildfire Vulnerability Prediction.

Input: Selected Principal Components Values (W') for some Event E
Output: Wildfire Vulnerability Class
1 **for** $i \leftarrow 1$ ***to*** 5 **do**
2 Calculate Posterior Probability $P\left(\frac{C_i}{W'}\right)$using Eq. (17) and (18)
3 Wildfire Vulnerability Class of event E = $\arg max_{\,i}\left(P\left(\frac{C_i}{W'}\right)\right)$
4 **return** C_i

hectares of forest land that would potentially be engulfed in wildfires would aid in efficient mitigation and management of wildfires. Therefore, burnt area prediction component employs support vector regression (SVR) to serve this purpose. The basic principle of SVR is to find a nonlinear mapping from input space to output space, also known as feature space, followed by the construction of a linear problem in this feature space. Consider $(a_i,\ b_i)_{i=1}^N$ as a training set, where $a_i \in \Re^p$ is the input vector comprising of WIPs, $b_i \in \Re$ denotes the corresponding output, i.e., the burnt area, and N denotes the number of instances in the training dataset. The linear problem is formulated as

$$\begin{aligned}\text{Burnt Area} = b_i = f(a,\ w) &= \sum_{i=1}^{m} w_i\, \varphi_i(a) + \beta \\ &= w^T \varphi(a) + \beta \qquad (19)\end{aligned}$$

where $w = [w_1,\ w_2, \ldots, w_m]$is the weight vector, β the bias, and $\varphi(a) = [\varphi_1(a),\ \varphi_2(a), \ldots, \varphi_m(a)]$ the nonlinear mapping according to kernel function $K(a,\ a_i) =\ \varphi^T(a)\ \varphi(a_i)$. The radial basis function, a widely used kernel function, is adopted in the proposed system and is defined as

$$K(a, a_i) = \exp\left(-\gamma\, ||a -\ a_i||^2\right),\ \gamma > 0 \qquad (20)$$

where γ is the kernel parameter. The accuracy of this component is assessed in terms of root-mean-squared error (RMSE) and mean absolute error (MAE). The results of this component would assist the forest department and disaster response agencies to deal with wildfires efficiently.

2) Wildfire Forecasting Layer: The wildfire forecasting layer is responsible for the efficient and reliable estimation of future wildfire incidences by using the analysis of historical along with current wildfire phenomena and observed meteorological attributes. To accomplish this task, state-of-the-art time series analysis and forecasting model, known as SARIMA, is employed in the proposed system. SARIMA, also known as seasonal ARIMA, is a version of the ARIMA model, which, in turn, encompasses auto-regressive (AR) and moving average (MA) models, each of which is described ahead. In the AR model, the future value of a variable is assumed to be a linear combination of a finite number of past observations and a random error together with a constant term. The *i*th order of the AR model denoted as *AR(i)* is mathematically expressed as

$$\begin{aligned}\mathbb{V}_t &= \mathrm{c} + \sum_{k=1}^{i} \phi_k \mathbb{V}_{t-k} + \varepsilon_\mathrm{t} \\ &= \mathrm{c} + \phi_1 \mathbb{V}_{t-1} + \phi_2 \mathbb{V}_{t-2} + \cdots + \phi_i \mathbb{V}_{t-i} + \varepsilon_\mathrm{t} \qquad (21)\end{aligned}$$

where c is the constant term, $\mathbb{V}_{t-1},\ \mathbb{V}_{t-2} \ldots \mathbb{V}_{t-i}$ the past wildfire vulnerability values, ϕ the auto regression coefficient, and ε_t the random error or random shock at time *t*. In the MA model, the future value of a variable is assumed to be a linear combination of a finite number of past forecast errors. The *j*th order of the MA model denoted as *MA(j)* is mathematically denoted as

$$\begin{aligned}\mathbb{V}_t &= \mu + \sum_{l=1}^{j} \theta_\mathrm{l}\, \varepsilon_{\mathrm{t-l}} + \varepsilon_\mathrm{t} \\ &= \mu + \theta_1\, \varepsilon_{\mathrm{t-1}} + \theta_2\, \varepsilon_{\mathrm{t-2}} + \cdots + \theta_\mathrm{j}\, \varepsilon_{\mathrm{t-j}} + \varepsilon_\mathrm{t} \qquad (22)\end{aligned}$$

where μ is the mean of the series, $\varepsilon_{\mathrm{t-1}}, \varepsilon_{\mathrm{t-2}} \ldots \varepsilon_{\mathrm{t-j}}$ the forecast errors noted in the past wildfire vulnerability values, θ the MA coefficient, and ε_t the random error or random shock at time *t.* ARIMA, which encompasses both AR and MA models, is denoted as *ARIMA(i, j)*, where *i* represents the order of the auto-regression part and *j* the order of the MA part. It is mathematically expressed as

$$\mathbb{V}_t = \mathrm{c} + \varepsilon_\mathrm{t} + \sum_{k=1}^{i} \phi_k \mathbb{V}_{t-k} + \sum_{l=1}^{j} \theta_\mathrm{l}\, \varepsilon_{\mathrm{t-l}}. \qquad (23)$$

Thus, forecasting depends on the past values of the variable under consideration, as well as the past values of forecast errors. Since climate plays an important role in triggering wildfires, and climate follows a seasonal or annual cycle, it becomes essential to include seasonal components in traditional ARIMA model. Such a model, called seasonal ARIMA or SARIMA, is abbreviated as SARIMA $(i,\ d,\ j)\,(I,\ D,\ J)\,N$, where lowercase letters *i, d, j* denote nonseasonal AR order, differencing, and MA order; uppercase letters *I, D, J* denote seasonal AR order, differencing, and MA order. *N* denotes the number of periods per season. The backshift notation for SARIMA is expressed as

$$\phi_{\text{AR}}(B)\, \phi_{\text{SAR}}\left(B^N\right)\left(1 - B^N\right)^D \mathbb{V}_t = \theta_{\mathrm{MA}}(B)\, \theta_{\mathrm{SMA}}\left(B^N\right) \varepsilon_\mathrm{t} \qquad (24)$$

where ϕ_{AR} is the nonseasonal auto-regression coefficient, ϕ_{SAR} the seasonal auto-regression coefficient, θ_{MA} the nonseasonal MA coefficient, θ_{SMA} the seasonal MA coefficient, and *B* the backshift operator.

The forecasting of wildfire vulnerability level $\mathbb{V}_t$ includes four sequential steps, namely, 1) model specification; 2) parameter estimation; 3) diagnostic checking; and 4) forecasting. Algorithm 3 illustrates the working of the wildfire forecasting layer. Diagnostic checking involves the performance evaluation of different models based on four measures of accuracy, namely, 1) MAE; 2) mean square error (MSE); and RMSE . The model with the minimum values of MAE, MSE, and RMSE is selected for forecasting.

3) Cloud Storage Layer: This component is responsible for storing wildfire oriented data, as well as compiling the analysis

Algorithm 3: Wildfire Vulnerability Forecasting.

Input: Wildfire Vulnerability Dataset
Output: Forecasted Wildfire Vulnerability Level for time *t*

1: Determine the values of *d*, *D,* and *N* to make data stationary. $d \leftarrow 1$ for nonseasonal part and $D \leftarrow 1$ and $N \leftarrow 12$ for seasonal part.
2: Determine the best combination order of SARIMA, i.e., the values of *i*, *j*, *I,* and *J* by examining auto-correlation function (ACF) and partial ACF (PACF) plots.
3: Determine auto-regression and MA coefficients for both seasonal and nonseasonal parts.
4: Choose the most appropriate model based on the smallest values of MAE, MSE, and RMSE.
5: The wildfire susceptibility level for time *t* is forecasted using (24).
6: **return** $\mathbb{V}_t$

results regarding wildfire outbreak, the susceptibility of various forest terrains to fire, fire-prone seasons, etc. This would aid the forest department and disaster-management agencies in efficiently fighting small and large wildfires by the creation of new fire lines, maintenance of existing ones, creation of buffer zones, clearance of dry fuels from the ground, etc.

IV. Implementation Analysis

This section provides insights into experimental setup and the subsequent performance analysis of the proposed system. This section comprises of six sub-sections, namely, 1) data acquisition; 2) ANOVA and Tukey's *post hoc* aided energy conservation analysis; 3) PCA centric dimensionality reduction analysis; 4) naïve Bayes assisted wildfire vulnerability prediction analysis; 5) SVR centric burnt area prediction; and 6) SARIMA assisted wildfire vulnerability forecasting analysis.

A. Data Acquisition

Wildfire oriented datasets of four forest divisions of Punjab, India, i.e., Pathankot, Hoshiarpur, Dasuya, and Garhshankar, are obtained from the Punjab Forest Department [12]. Also, the data are obtained from the Forest Survey of India [13]. The acquired data are stored at the Amazon Elastic Compute Cloud (Amazon EC2). The proposed system is assessed using the Amazon EC2 t2.medium instance with two virtual CPUs and 4 GB memory.

B. ANOVA and Tukey's Post Hoc Aided Energy Conservation Analysis

The performance of the proposed energy conserving mechanism is gauged by the actual deployment of four periodic sensor nodes, one in each of the selected forest divisions. Each forest division encompasses forest ranges, which further include forest blocks where the sensor nodes are deployed. These sensor nodes are linked to the Android-based mobile phones of forest guards, which serve as e-nodes. A Javascript-based application is installed on the mobile phones to execute the energy-conserving algorithm. The server situated in the block office serves as a gateway that relays the captured data to the cloud.

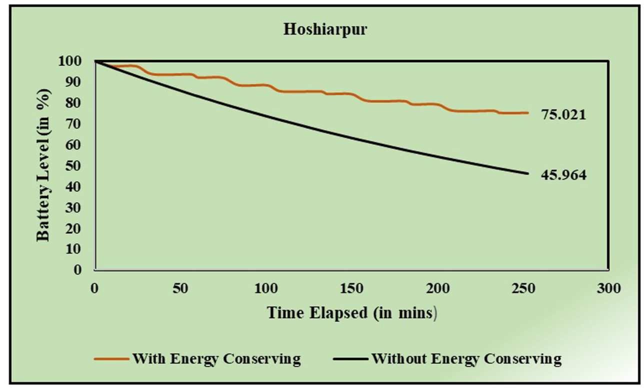


Fig. 3. Energy conservation analysis—battery level of sensor node with and without energy conserving mechanism.

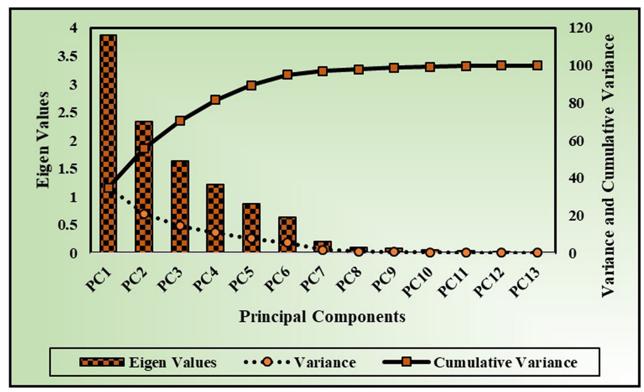


Fig. 4. Scree plot of PCA.

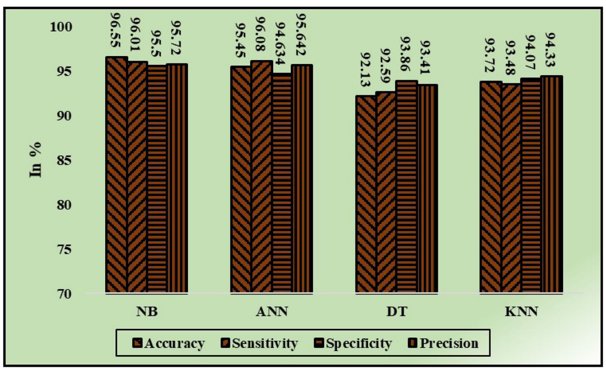
The battery level of the sensor nodes with and without energy conserving mechanism is considered as a criterion for evaluating its efficacy. The initial values of active and sleep intervals, i.e., τ_a and τ_s, in the monitored forest divisions are set, considering the proneness of each division to wildfires. Since the Hoshiarpur division is highly vulnerable to wildfires, followed by Pathankot, Garhshankar, and Dasuya, therefore, $\tau_a^H > \tau_a^P > \tau_a^G > \tau_a^D$, which infers $\tau_s^H < \tau_s^P < \tau_s^G < \tau_s^D$, where τ_a^Z and τ_s^Z denote the active and sleep intervals of the sensor nodes in any Z forest division. Implementation results reveal significant battery savings owing to the proposed energy conserving mechanism, the results of which are depicted in Fig. 3. The graph illustrates the battery level of the deployed sensor node with and without the proposed energy conserving mechanism in the Hoshiarpur forest division. For the Pathankot, Garhshankar, and Dasuya forest divisions, the battery percentages recorded without energy conserving mechanism for the considered period are 41%, 36%, and 32%, respectively. The battery percentages recorded with energy saving mechanism are approximately 79%, 83%, and 88%.

C. PCA Centric Dimensionality Reduction Analysis

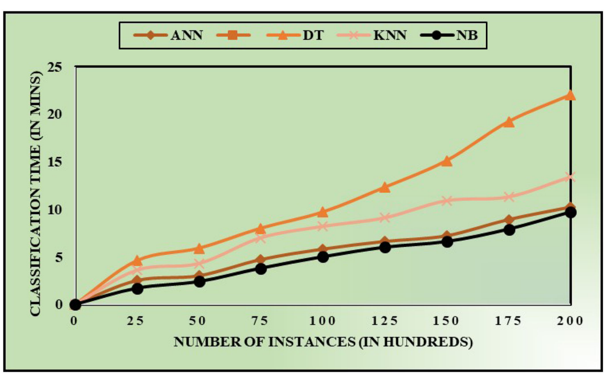
The proposed framework employs PCA for reducing the dimensionality of the WIPs. The PCA is implemented in R studio using the princomp() function, the results of which are shown in

TABLE II
CLASSWISE ACCURACY EVALUATION OF NB

Class	Accuracy	Sensitivity	Specificity	Precision
NR	0.96	0.9658	0.9644	0.94118
LR	0.9775	0.9425	0.94937	0.959
MR	0.97	0.955	0.95324	0.9642
HR	0.9525	0.9648	0.94382	0.9677
EHR	0.9675	0.9725	0.96429	0.954
Mean	0.9655	0.96012	0.955024	0.957216



(a)



(b)

Fig. 5. Comparative analysis of NB with other state-of-the-art classifiers. (a) Performance metrics. (b) Classification time.

Fig. 4. The first five PCs have eigenvalues >1 and explain the cumulative variance of 89.4%. Therefore, the first five PCs projecting the majority of information are selected and relayed to the naïve Bayes classifier for wildfire vulnerability prediction.

D. Naïve Bayes Assisted Wildfire Vulnerability Prediction Analysis

The wildfire vulnerability of a forest terrain is predicted using the naïve Bayes classifier. The monitored forest terrain is assigned one of the five vulnerability classes from no risk (1) to extremely high risk (5). Statistical measures namely, accuracy, sensitivity, specificity, and precision are used as the criteria for assessing the performance of the naïve Bayes classifier. The values of the above-mentioned statistical measures for all the five wildfire vulnerability classes are given in Table II.

Also, the performance of the naïve Bayes is compared with various state-of-the-art classifiers, namely, artificial neural networks, k-nearest neighbor, decision tree, and support vector machine, the results of which are shown in Fig. 5(a) and (b). Based on the results, it is concluded that the naïve Bayes outperforms other classification algorithms, thereby making it an appropriate choice for the considered scenario.

TABLE III
ACCURACY EVALUATION OF SVR

Division	Hoshiarpur	Pathankot	Garhshankar	Dasuya
MAE	2.242	1.400833	2.079167	1.991667
RMSE	7.104392	4.602249	6.435906	6.032715

TABLE IV
ACCURACY EVALUATION OF SARIMA

	MAE	RMSE	MSE
Hoshiarpur	0.083333	0.288675	0.083333
Pathankot	0.166667	0.408248	0.166667
Garhshankar	0.083333	0.288675	0.083333
Dasuya	0.166667	0.408248	0.166667

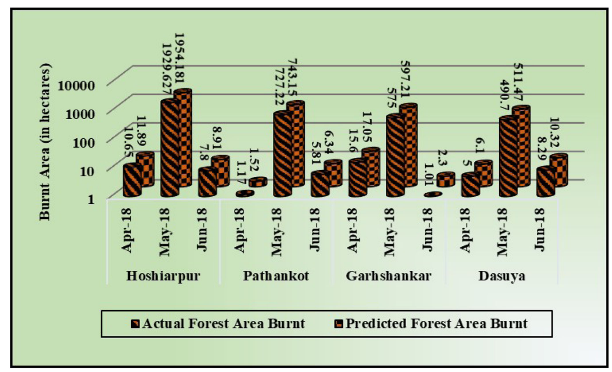


Fig. 6. Burnt area prediction.

E. SVR Centric Burnt Area Prediction Analysis

The proposed system employs SVR for estimating the area of forest that could potentially be engulfed by wildfire incidences. SVR is implemented using the svm() function in R studio. Fig. 6 depicts the results of SVR in all the considered forest divisions. Table III lists the accuracy measures, i.e., MAE and RMSE, recorded in all the four divisions.

F. SARIMA-Assisted Wildfire Vulnerability Forecasting Analysis

The wildfire vulnerability level is forecasted with the help of the SARIMA model by using the results produced by the naïve Bayes classifier. SARIMA model is implemented using R Studio on Amazon EC2. Sarima—an R package containing diverse functions, classes, and methods for time series modeling with ARIMA and associated models—is installed. The best combination order of SARIMA for each forest division is chosen through the examination of ACF and PACF plots. Amongst all the plausible models, SARIMA (0, 1, 1) (0, 1, 1)12 is chosen as the best fit model for both the Hoshiarpur and Garhshankar divisions; SARIMA (1, 1, 0)(0, 1, 1)12 is selected as the best fit model for the Pathankot division; SARIMA (1, 1, 0)(1, 1, 0)12 is selected as the best fit model for the Dasuya division. The wildfire vulnerability forecasting results are depicted in Fig. 7(a)–(d). In addition, Table IV represents the

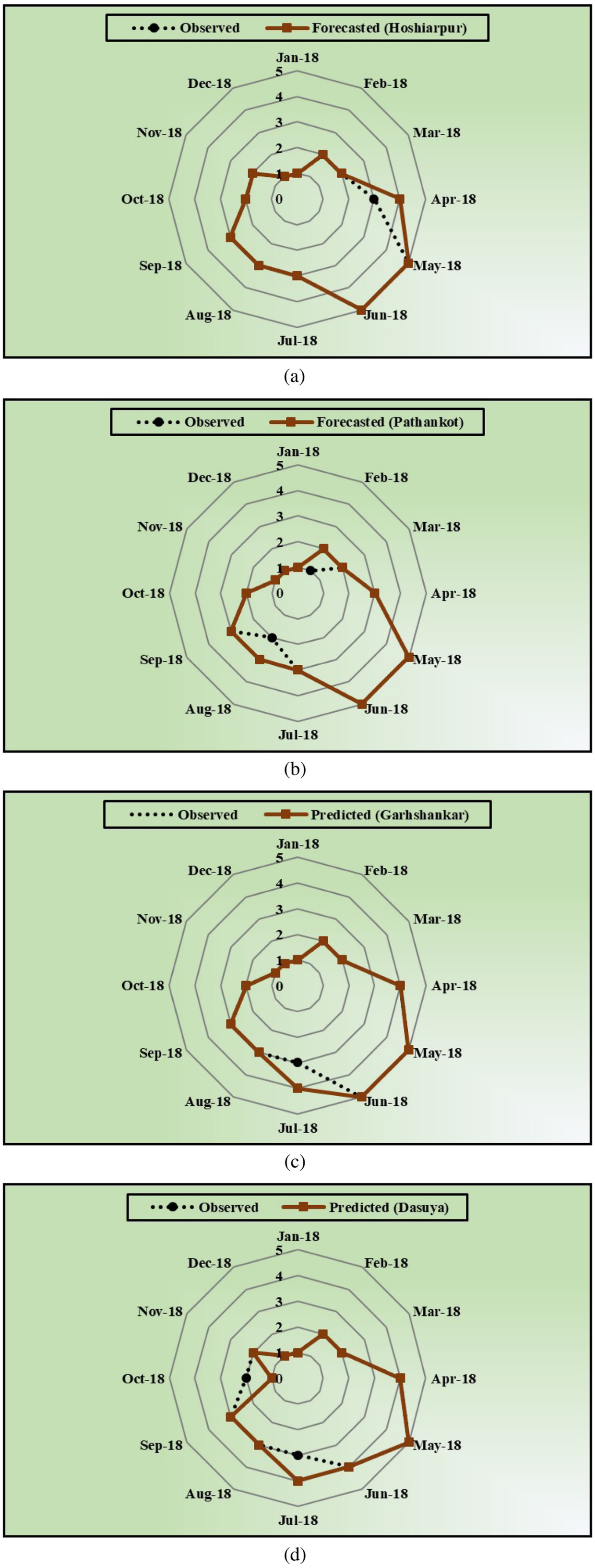


Fig. 7. Wildfire susceptibility forecasting using seasonal ARIMA (a) Hoshiarpur division. (b) Pathankot division. (c) Garhshankar division. (d) Dasuya division.

forecast accuracy evaluation results in terms of MAE, MSE, and RMSE.

V. Conclusion

In this paper, fog-assisted, IoT-enabled energy-efficient layered framework is proposed for wildfire prediction and forecasting. Initially, IoT sensors are deployed to sample the forest block for WIPs. The key feature of the proposed framework is an efficient adaptive method of data sampling based on data variability, thus increasing the lifetime of resource-constrained sensors. Furthermore, network bandwidth is optimally utilized by employing PCA for the dimensionality reduction of the WIPs. The reduced set of the WIPs is analyzed at Cloud Layer to predict the wildfire vulnerability level of a forest terrain using NB classifier. In addition, forest area (in hectares) that could possibly be burnt in case of wildfire outbreak is also predicted using SVR. The proposed architecture is implemented on Amazon EC2 t2.medium instance. The obtained results prove that the model is highly efficient in wildfire prediction and forecasting. The analysis results of the system serve as valuable inputs for the forest department and other disaster-management agencies to efficiently mitigate and manage forest fires.

References

[1] Facts + statistics : Wildfires. [Online]. Available: https://www.iii.org/fact-statistic/facts-statistics-wildfires. Accessed: Jun. 10, 2019.

[2] S. K. Sood and I. Mahajan, “A fog based healthcare framework for Chikungunya,” *IEEE Internet Things J.*, vol. 5, no. 2, pp. 794–801, Apr. 2018.

[3] P. Verma and S. K. Sood, “Fog-assisted IoT-enabled patient health monitoring in smart homes,” *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1789–1796, Jun. 2018.

[4] Y. E. Aslan, I. Korpeoglu, and O. Ulusoy, “A framework for use of wireless sensor networks in forest fire detection and monitoring,” *Comput., Environ. Urban Syst.*, vol. 36, no. 6, pp. 614–625, 2012.

[5] A. R. Ulucinar, I. Korpeoglu, and A. E. Cetin, “A Wi-Fi cluster based wireless sensor network application and deployment for wildfire detection,” *Int. J. Distrib. Sensor Netw.*, vol. 10, no. 10, pp. 1–12, 2014.

[6] A. Molina-Pico, D. Cuesta-Frau, A. Araujo, J. Alejandre, and A. Rozas, “Forest monitoring and wildland early fire detection by a hierarchical wireless sensor network,” *J. Sensors*, vol. 2016, 2016, Art. no. 8325845.

[7] S. Garcia-Jimenez, A. Jurio, M. Pagola, L. De Miguel, E. Barrenechea, and H. Sola, “Forest fire detection: A fuzzy system approach based on overlap indices,” *Appl. Soft Comput.*, vol. 52, pp. 834–842, 2017.

[8] H. Lin, X. Liu, X. Wang, and Y. Liu, “A fuzzy inference and big data analysis algorithm for the prediction of forest fire based on rechargeable wireless sensor networks,” *Sustain. Comput. Inform. Syst.*, vol. 18, pp. 101–111, 2018.

[9] J. Toledo-Castro, P. Caballero-Gil, N. Rodríguez-Pérez, I. Santos-González, C. Hernández-Goya, and R. Aguasca-Colomo, “Forest fire prevention, detection, and fighting based on fuzzy logic and wireless sensor networks,” *Complexity*, vol. 2018, 2018, Art. no. 1639715.

[10] N. Kaur and S. K. Sood, “An energy-efficient architecture for the Internet of Things (IoT),” *IEEE Syst. J.*, vol. 11 no. 2, pp. 796–805, Jun. 2017.

[11] I. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer-Verlag, 2011.

[12] Pb forests. [Online]. Available: http://pbforests.gov.in/. Accessed: Jun. 10, 2019.

[13] Forest Survey of India. [Online]. Available: https://fsi.nic.in. Accessed: Jun. 10, 2019.