# Cloud-Centric IoT-Based Green Framework for Smart Drought Prediction

Amandeep Kaur and Sandeep K. Sood

***Abstract*—Drought is a catastrophic natural disaster with significant impact on financial stability, hydrological budget, public health, and agricultural productivity. Numerous drought indices have been introduced to quantify the severity of droughts, but the majority of them are incapable of demonstrating the changes in crucial drought causing elements. Internet of Things (IoT) is appropriate to monitor time-critical environmental parameters. This article proposes an energy-efficient cloud-centric system to assess the drought for the current situation and predict for the future time frame. The architecture determines the active and sleep interval of IoT sensors based on the analysis of data variability using the Bartlett test. The dimensionality of the data about drought causing elements is reduced using kernel principal component analysis (KPCA) at the fog layer. The intensity of drought is determined at the cloud layer using naïve-Bayes classifier, and drought severity for different time periods is predicted using the seasonal autoregressive integrated moving average model. Experimentation and performance analysis prove the efficiency of the proposed system in assessing and predicting the drought with better correlations with drought-causing attributes. Furthermore, it shows significant energy savings as compared to other schemes.**

***Index Terms*—Cloud computing (CC), energy efficiency, fog computing, Internet of Things (IoT), kernel principal component analysis (KPCA), seasonal autoregressive integrated moving average (SARIMA) model.**

## I. Introduction

DROUGHT is a destructive natural disaster that causes enormous irrevocable damages. The incidence and scale of drought events are increasing due to human activities and global warming [1]. Inadequate precipitation for a long period and scarcity of surface water cause vegetation degradation, soil depletion, groundwater deficiency, etc. Declined agricultural production and poor public health deter economic growth. Economic collapse leads to civil upheaval and political turmoil. During 1997–2017, droughts have affected around 1.5 billion people, and in 2017, drought accounted for 1.5 million displacements [2]. Rise of drought conditions accentuates the pressing requirement of a system for timely assessment and prediction of drought. A number of drought indices, such as the standard precipitation index (SPI), the normalized drought vegetation index (NDVI), and the Palmer drought severity index (PDSI), are used to quantify the drought severity. However, each drought index measures different meteorological and climatic attributes. There is a lack of globally accepted index that can assess the drought intensity in any part of the world by considering a broad range of causing elements [3]. This highlights the need of the real-time system that can assess and predict drought correctly on the basis of a wide variety of causing variables. This can be attained by tapping the immense potential of cloud computing (CC), big data, Internet of Things (IoT), and fog computing [4], [5].

This article proposes a layered framework for drought assessment and prediction. Initially, the IoT sensors are placed for the continuous monitoring of area under study that provides a humongous amount of data. However, the limited battery power of sensor nodes makes it difficult to collect the environmental data for a prolonged time period. Apart from this, the replacement of the batteries especially in faraway lands is quite laborious. To tackle this obstacle, a mechanism to save the battery life to determine the active and sleep interval of the sensor node depending upon the variance of the received data has been employed in the fog layer. The intensity level of drought is assessed and predicted at the cloud layer by utilizing its enormous computing power. The objectives of the proposed system are: 1) data collection by deploying the IoT sensors in the region; 2) efficient-energy utilization of sensors to improve the sensor network lifetime; 3) efficient bandwidth utilization by reducing dimensions at the fog layer; 4) assessment and prediction of drought intensity for the considered area; and 5) transmission of predicted evaluations to drought management agencies.

This article is organized as follows. Section II investigates some of the literature on drought prediction and energy-efficient sleep scheduling schemes. Section III describes the proposed energy-efficient framework for drought assessment and prediction. Section IV reveals the experimental setup of the proposed system, results obtained, and its performance analysis. Finally, this article ends with conclusive remarks in Section V.

## II. Related Work

Many studies have predicted future situations and data successfully with higher accuracy [6]–[8]. Comparative analysis

Manuscript received September 6, 2019; revised October 1, 2019 and October 24, 2019; accepted October 30, 2019. Date of publication November 5, 2019; date of current version February 11, 2020. *(Corresponding author: Amandeep Kaur.)*

A. Kaur is with the Department of Computer Science and Engineering, Guru Nanak Dev University, Gurdaspur 143521, India (e-mail: ghotraaman17@gmail.com).

S. K. Sood is with the Department of Computer and Informatics, Central University of Himachal Pradesh, Dharamshala, India (e-mail: san1198@gmail.com).

Digital Object Identifier 10.1109/JIOT.2019.2951610

TABLE I
COMPARATIVE ANALYSIS OF RELATED WORK

| Authors | Contribution | Application | Input Variables | Smart Technology Used | Real-time Aspect | Energy Saving | Decision Making Model |
|---|---|---|---|---|---|---|---|
| Poornima et al.[6] | Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network | Long-term prediction of meteorological droughts | SPI, SPEI | No | No | No | Long Short-term Memory (LSTM), ARIMA. |
| Soh et al. [7] | Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia | Short-term and Long-term drought forecasting | SPEI | No | No | No | Wavelet-ARIMA-ANN, Wavelet-ANFIS |
| Ali et al. [8] | Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting | Drought forecasting using climatic variables | Rainfall, Temperature, Humidity, Southern Oscillation Index | No | No | No | Committee Extreme Learning Machine |
| Yu et al. [9] | Investigation of drought-vulnerable regions in North Korea using remote sensing and cloud computing climate data | Drought Assessment to identify drought vulnerable regions | Precipitation, NDVI, Land cover | Remote Sensing, CC | No | No | No |
| Zou et al. [10] | MapReduce functions to remote sensing distributed data processing-Global vegetation drought monitoring as example | Cloud framework for drought monitoring | Vegetation | CC | Yes | No | MapReduce Paradigm |
| Du et al. [11] | Monitoring System for Wheat Meteorological Disasters using Wireless Sensor Networks | Diagnosis of Meteorological droughts, Real-time monitoring | Temperature, Humidity, Soil Temperature, Soil Moisture, Solar Radiation Intensity, Precipitation, Wind Speed | Wireless Sensor N/w | Yes | No | Threshold-based method |
| Demisse et al. [12] | Information Mining from Heterogeneous Data Sources: A Case Study on Drought Predictions | Drought prediction using data from different sources | SDNDVI, Ecological region, Soil water holding, Land use, SPI, Oscillation Index | No | No | No | Classification and Regression Tree |
| Hao et al. [13] | An integrated package for drought monitoring, prediction and analysis to aid drought modeling and assessment | Drought prediction by calculating univariate and multivariate drought indices | Precipitation, Run-off, Soil Moisture, Groundwater | No | No | No | Joint distribution & Kendell distribution function |
| Meca & Pech [14] | Forecasting SPEI and SPI Drought Indices Using the Integrated Artificial Neural Networks | Drought forecast, Comparison of different ANN models | SPI, SPEI | No | No | No | Integrated ANN |
| Proposed System | Cloud-Centric IoT based Green Framework for Smart Drought Prediction | Drought assessment, Drought prediction for different time frames | Groundwater, Streamflow, Soil moisture and Soil temperature at various levels, Temperature, Humidity, Precipitation, Evapotranspiration, Season | Fog Computing, CC,IoT | Yes | Yes | KPCA, Naive Bayes, SARIMA |

of the relevant and important work done by the authors in the drought prediction field on the basis of seven parameters is shown in Table I. This analysis reveals that most of the work has been done on the prediction of drought indices [9], [10], [17]. Most of the authors have considered a limited set of input attributes without any real-time aspect. Du *et al.* [14] have taken a lot of input variables into account and used the wireless sensor network but the network is utilized for monitoring purposes using a simple threshold-based method.

Many energy-efficient sleep scheduling schemes have been presented. Zhang *et al.* [18] in 2019 proposed a three-level sleep scheduling approach in which minimum active sensors are selected in such a way that all the inactive sensors have at least one active neighbor. The second and third levels deal with the discovery of active sensor groups and selecting one of them to be active in a particular time frame. Sakya and Sharma [19] in 2019 presented an energy-efficient adaptive MAE protocol to determine the duty cycle of sensors using regression by considering the number of packets waiting and remaining energy. Gupta *et al.* [20] in 2015 put forth a system that decides the next state of a sensor by examining its redundancy, communication, and sensing range of the neighboring heterogeneous nodes without producing any coverage hole. Mostafaci *et al.* [21] in 2016 proposed the use

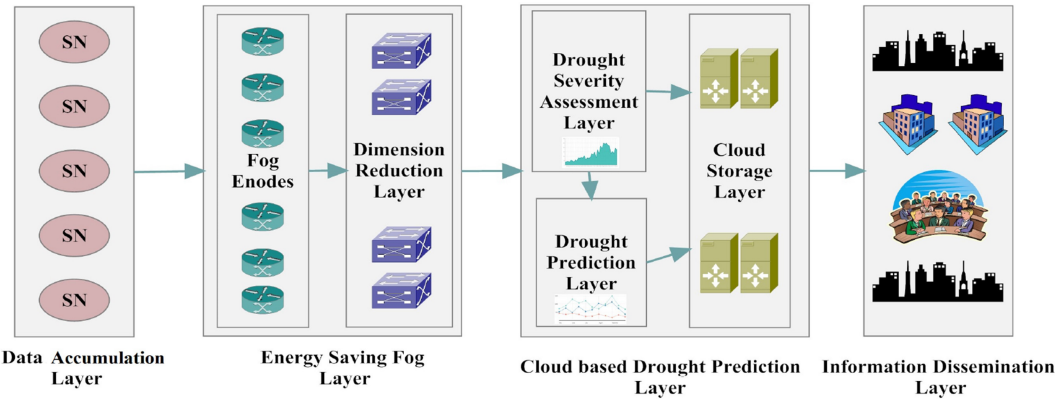


Fig. 1. Layered architecture of the proposed system.

of the probabilistic framework of leaving automata (PCLA) to determine sleep or active state of the minimum number of sensors that cover a connected region of interest. Liu *et al.* [22] in 2016 proposed the logical-correlation-based sleep scheduling mechanism (LCSSM) to determine the logical correlations between the events of different sensors. The corresponding sensors are activated or deactivated taking into account the valuable events they are performing and their logical correlation with other sensor nodes. Sahoo *et al.* [23] in 2017 put forward the use of the prescheduled k-coverage group scheduling algorithm to identify the groups of wanted sensors with the least overlapping and self-organized k-coverage scheduling algorithm to figure out the state of the sensor node considering the selected intersection points. Singh and Lobiyal [24] in 2013 proposed the use of Gaussian distribution to deploy the sensor nodes to avoid the energy hole near the sink node. Moreover, the sleep probability of the nodes is computed considering the sensor density and traffic.

Zebbane *et al.* [25] in 2015 introduced the group-based energy conserving protocol (GECP) to increase the life span of the wireless sensor network. First, the system finds the neighbors of all the nodes and then forms the groups in which the sensors with the same set of adjacent nodes belong to the same group. Then, one active node per group is chosen to be active at one time using the scheduling strategy. Oladimeji *et al.* [26] in 2017 proposed the stochastic selection of inactive nodes (SSINs) protocol to select the nodes to put into the sleep mode which affects coverage minimally. The authors used the genetically inspired method to select the cluster heads to get a balanced energy usage over the whole network area. Bagci and Yazici [27] in 2013 presented the energy-aware unequal clustering with fuzzy (EAUCF) that utilizes distance from the base station and remaining battery power as the fuzzy variables to compute competition radius of tentative heads of uneven clusters.

The aforementioned literature discussed the approaches to determine the active or sleep mode of the sensor nodes by taking factors like network coverage, residual power, distance from base station, etc., into account to prolong the sensor as well as network lifetime. None of the above authors have paid any attention toward the accuracy of collected data. The main contribution of this article is to determine the duration of the active and sleep interval of a sensor node considering the variance of the collected data with the aim of getting more accurate results by sensing maximum variance of drought-causing attributes per unit of an active time period.

## III. Proposed System

A fog-based energy-efficient framework for drought assessment and prediction of drought is depicted in Fig. 1. The proposed system consists of four layers: 1) data accumulation layer; 2) energy-saving fog layer; 3) cloud-based drought prediction layer; and 4) information dissemination layer. The data accumulation layer deals with the deployment of sensor nodes in the region to accumulate raw data on causing variables. The energy-saving fog layer analyzes the data for the effective use of limited sensor energy and reduces the dimensions of collected data. The cloud-based drought prediction layer evaluates the intensity level of drought and predicts the drought conditions for future time frame to share the results with the information dissemination layer.

### A. Data Accumulation Layer

An effective drought assessment framework depends on the accuracy of the collected information. This layer acquires the raw data of climatic and environmental drought causing attributes as explained in Table II. The hydrological variables present the current hydrological state, and the climate and environmental variables define the meteorological state of the study area. To deploy sensors, the study area is divided into hexagons and the sensor node is placed in the center of each hexagon. The size of the hexagon depends upon the range of sensors used.

### B. Energy-Saving Fog Layer

*1) Energy-Saving Layer:* The fog layer receives the raw data from the data accumulation layer and sends it to the dimension reduction layer. The deployed sensor nodes are battery powered, and there is need of an energy-efficient data collection approach for power-constrained sensors. For this purpose, a mechanism to calculate the active and sleep time depending upon the nature of sensor data has been proposed in this architecture. The variance of the sensed data is analyzed

TABLE II
DROUGHT-INDUCING VARIABLES

| Hydrological Variables | | | |
|---|---|---|---|
| S. No | Variable | Description | Sensors |
| 1. | Groundwater | Upper waterline below the earth surface. | Piezometers, Groundwater level, Water flow, Bubbler and Soil moisture sensors, Tensiometers |
| 2. | Soil Moisture | Water content in the soil at 2, 4, 8, 20 and 40 inches below the surface. | |
| 3. | Streamflow | Water volume present in the region at that time. | |
| Climate and Environmental Variables | | | |
| S. No | Variable | Description | Sensors |
| 1. | Air Temperature | Atmospheric temperature of the region. | Non-Contacting and Contacting Temperature Sensors, Humidity Sensors, Rain Gauges, Precipitation Sensors, Evapotranspiration Sensors |
| 2. | Humidity | Amount of vapours present in the atmosphere. | |
| 3. | Soil Temperature | Measure of heat in the soil at the depth of 2, 4, 8, 20 and 40 inches. | |
| 4. | Evapotranspiration | Amount of vaporisation due to transpiration and evaporation. | |
| 5. | Precipitation | Water that falls on the region ground. | |
| 6. | Season | Meteorological Seasons: Spring, Summer, Autumn, Monsoon and Winter. | |

TABLE III
NOTATION TABLE FOR THE FOG LAYER

| Notation | Description |
|---|---|
| $n_h$ | Measures taken by sensor during $h^{th}$ active period |
| $\sigma_h^2$ | Variance of $h^{th}$ active period |
| N | Measures taken by the sensor during previous H active periods |
| $\sigma^2$ | Pooled Variance |
| S | Bartlett Test Static |
| $T_{th}$ | Threshold value |
| $T_a$ | Initial Active time period |
| $T_s$ | Initial Sleep time period |
| $AT_\beta^{\alpha+1}$ | $\alpha^{th}$ Active time period of $\beta^{th}$ sensor node. |
| $ST_\beta^{\alpha+1}$ | $\alpha^{th}$ Sleep time period of $\beta^{th}$ sensor node. |
| $A_{Q*Z}$ | Data set A with Q number of observations and Z drought causing attributes |
| $a_q$ | $q^{th}$ observation in A |
| $\Psi$ | Non-linear mapping function |
| $\widetilde{F}$ | High dimensional space with W dimensions |
| $R^Z$ | Z dimensional feature space |
| CoV | Covariance matrix of $\widetilde{F}$ space |
| K | Kernel Matrix of Q*Q matrix |
| $\widehat{K}$ | Centralized Kernel Matrix |
| $\lambda$ | Eigenvalue of covariance and kernel matrix |
| $\chi$ | Eigenvector of covariance matrix |
| $\chi^m$ | $m^{th}$ eigenvector of covariance matrix |
| $\gamma$ | Eigenvector of kernel matrix |
| $b_s$ | M dimensional KPCA projection of input $a_s$ |

using the Bartlett test at enode to avoid redundant information and sensor active and sleep intervals are updated accordingly. It will increase the lifetime of the sensor nodes, and only significant information will be forwarded to the cloud layer.

The Bartlett test is used to check whether data from different samples have equal variances. The samples are taken from the previous active periods of a sensor. If the variances are not significant, the next active time period of the sensor is decreased and the sleep time period is increased. If there are $H$ samples taken from the previous $H$ active periods of a sensor, the Bartlett test statistic ($S$) is calculated as follows:

$$S = \frac{(N-H)\ln\left(\sigma^2\right) \ - \sum_{h=1}^{H}(n_h-1)\ \ln\left(\sigma_h^2\right)}{\lambda} \tag{1}$$

where $n_h$ is the measurement taken by the sensor during the $h$th active period, and $\sigma_h^2$ is the variance for that period

$$N = \sum_{h=1}^{H} n_h \tag{2}$$

$$\lambda = 1 + \frac{1}{3(H-1)}\left(\sum_{h=1}^{H}\left(\frac{1}{n_h-1}\right) - \frac{1}{N-H}\right) \tag{3}$$

where $N$ is the measurement taken by the sensor during the previous $H$ active periods, and $\sigma^2$ is the pooled variance calculated as

$$\sigma^2 = \frac{1}{N-H}\sum_{h=1}^{H}\left(\sigma_h^2\right). \tag{4}$$

If $S$ is greater than threshold ($T_{\text{th}}$), the next active and sleep time interval is set to constants $T_a$ and $T_s$, respectively. If $S$ is less than $T_{\text{th}}$, the active interval is decreased and the sleep interval is increased as follows:

$$AT_\beta^{\alpha+1} = T_a\left(1 - \left(1 - \frac{S}{T_{\text{th}}}\right)\right) \tag{5}$$

$$ST_\beta^{\alpha+1} = T_s\left(1 + \left(1 - \frac{S}{T_{\text{th}}}\right)\right) \tag{6}$$

where $AT_\beta^\alpha$ = $\alpha$th active time period of the $\beta$th sensor node, and $ST_\beta^\alpha$ = $\alpha$th sleep time period of the $\beta$th sensor node. The value of $T_{\text{th}}$ is determined from a chi-square table by considering the values of $H$ and given significance level (SL). Algorithm 1 explains the complete procedure step by step to save energy at the fog layer and Table III explains all the notations. In the proposed architecture, samples from the previous two active intervals are considered to determine the value of $S$.

*2) Dimension Reduction Layer:* The analysis of such a vast amount of collected raw data is a challenging task. Dimension reduction is used to get a reduced data set with all the critical information without affecting the analytical results. The proposed framework applies kernel principal component analysis (KPCA) for this purpose. KPCA is utilized to extract the nonlinear characteristics of real-time environmental, climatic, and water supply attributes. The input data are projected over a linearly separated high-dimensional space using the kernel method and then principal component analysis (PCA) is performed to extract the principal components (PCs)

**Algorithm 1** Energy Saving at Enode

Input: Samples from previous two active periods of sensors.
Output: Next sleep and active interval for sensors.
Step 1.For every sensor node ($\beta$)
  Step 1.1 If (mode$_\beta$="active")
    Step 1.1.1 Calculate S as per equation(1) by taking measures for previous two active time periods $AT_\beta{}^{\alpha-1}$, $AT_\beta{}^{\alpha}$ of $\beta^{th}$ sensor.
    Step 1.1.2 If(S $>T_{th}$)
      Step 1.1.2.1 $AT_\beta{}^{\alpha+1} = T_a$
      Step 1.1.2.2 $ST_\beta{}^{\alpha+1} = T_s$
    Step 1.1.3 Else
      Step 1.1.3.1The next active period $AT_\beta{}^{\alpha+1}$ and next sleep interval period $ST_\beta{}^{\alpha+1}$ is calculated using equations (5) and (6) respectively.
    Step 1.1.4 End if
  Step 1.2 End if
  Step 1.3 If (mode$_\beta$="sleep")
    Step 1.3.1Send a wakeup signal after $ST_\beta{}^{\alpha}$ time interval.
Step 2. Exit.

**Algorithm 2** Dimension Reduction Using KPCA

Input: Dataset $A_{Q*Z}$ with Q number of tuples and Z drought inducing variables
Step 1. Calculate Q*Q kernel matrix K.
Step 2. Calculate Centralized kernel Matrix K using equation (14).
Step 3. Calculate eigenvectors $\gamma$ for the centralized kernel matrix.
Step 4. Sort eigenvectors in descending order of eigenvalues and select highest M of them.
Step 5. Normalize eigenvectors using equation (15).
Step 6. For any $a_{new} \epsilon$ A, transformed $b_{new}$ is calculated using equation (16).

of the extended space, ultimately realizing the original space.

We assume $A = [a_1, a_2, a_3, \ldots, a_Q]$ is the observation in the data set $A_{Q*Z}$ , where $a_q = [a_{q1}, a_{q2}, \ldots, a_{qZ}]$. The data are mapped to a high-dimensional vector space $\widetilde{F}$ with dimensions $W$ using a nonlinear mapping $\Psi$

$$\Psi : a_q \epsilon R^Z \quad \rightarrow \quad \Psi(a_q) \epsilon \ \widetilde{F}$$

where $W \gg Q$. The covariance matrix of $\widetilde{F}$ space can be calculated by

$$CoV = \frac{1}{Q}\sum_{q=1}^{Q}\left(\Psi(a_q) - \text{mean}\right)\left(\Psi(a_q)^T - \text{mean}\right) \quad (7)$$

where mean $= \sum_{q=1}^{Q} \Psi(a_q)$. Assume that the data set of $\widetilde{F}$ space is centralized, i.e., mean = 0. The PCs of feature space $[\Psi(a_1), \Psi(a_2), \Psi(a_3), \ldots, \Psi(a_q)]$ are determined by solving the following equation:

$$CoV \ \chi = \lambda\chi \quad (8)$$

where $\chi = [\chi^1, \chi^2, \ldots, \chi^Q]$ is eigenvector and $\lambda$ is eigenvalue of covariance matrix CoV. There exist coefficients $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_Q$ such that the orthogonal eigenvector $\chi$ is a linear combination of $[\Psi(a_1), \Psi(a_2), \Psi(a_3), \ldots, \Psi(a_Q)]$ and hence satisfies

$$\chi = \sum_{r=1}^{Q} \gamma_r \ \Psi(a_r). \quad (9)$$

Substituting (7) and (9) into (8) and multiplying with $\Psi(a_t)^T$ on both sides, it becomes

$$\frac{1}{Q}\Psi(a_t)^T \sum_{q=1}^{Q} \Psi(a_q) \ \Psi(a_q)^T \sum_{r=1}^{Q} \gamma_r \ \Psi(a_r)$$
$$= \lambda\Psi(a_t)^T \sum_{r=1}^{Q} \gamma_r \ \Psi(a_r). \quad (10)$$

Let us define a kernel matrix $K$ of size $Q * Q$ with its elements $K_{jl} = \Psi(a_j)^T\Psi(a_l) = k(a_j, a_l)$ as

$$\frac{1}{Q}\sum_{t=1,q=1}^{Q} k(a_t, a_q) \sum_{q=1,r=1}^{Q} \gamma_r \, k(a_q, a_r) = \lambda \sum_{t=1,r=1}^{Q} \gamma_r \, k(a_t, a_r). \quad (11)$$

The equation can be reformed as

$$K^2\gamma = \lambda QK\gamma. \quad (12)$$

The eigenvalue $\lambda$ and the corresponding eigenvector $\gamma$ can be obtained by solving the following eigenvector problem:

$$K\gamma = \lambda Q\gamma. \quad (13)$$

As aforementioned, it was assumed that the data in the high-dimensional space are centralized but generally this is not the case. Therefore, the data are centralized by calculating the centralized kernel matrix with its elements $\widehat{K}_{jl} = \widehat{\Psi}(a_j)^T\widehat{\Psi}(a_l)$, where $\widehat{\Psi}(a_j) = \Psi(a_j) - \text{mean}$. By expanding, we have

$$\widehat{K} = K - 1_Q \ K - K1_Q \ + 1_Q \ K1_Q \quad (14)$$

where $\widehat{K}$ is the centralized kernel matrix and matrix $1_Q = (1/Q)_{Q*Q}$ . Suppose $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M (M \leq Q)$ are top-$M$ eigenvalues of the kernel matrix. To ensure the normality of the PCs, the eigenvector $\gamma$ is rescaled as

$$\widehat{\gamma} = \frac{\gamma}{\sqrt{\lambda}\sqrt{Q}}. \quad (15)$$

For a new sample $a_s$, $M$-dimensional KPCA projection $b_s = [b_{s1}, b_{s2}, \ldots, b_{sM}]$ is

$$b_{sm} = \Psi(a_s)^T\chi^m \qquad \forall m = 1, 2, \ldots, M$$
$$= \Psi(a_s)^T \sum_{q=1}^{Q} \gamma_q^m \Psi(a_q) = \sum_{q=1}^{Q} \gamma_q^m k(a_q, a_s) \quad (16)$$

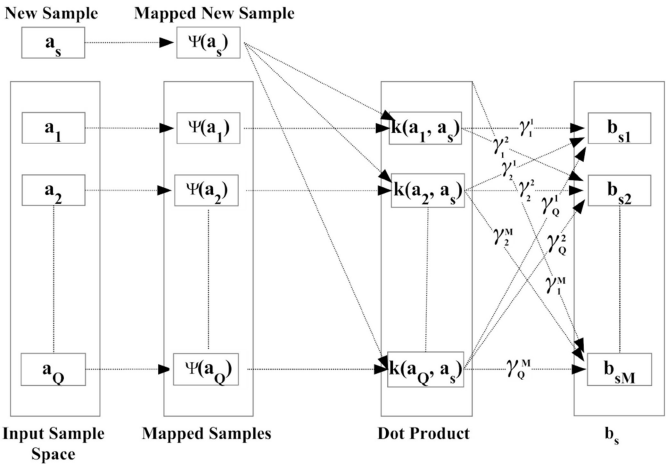


Fig. 2. Dimension reduction using KPCA.

where $\gamma_q^m$ is the $q$th element of the $m$th eigenvector of the kernel matrix $K$. Fig. 2 depicts the whole process.

### C. Cloud-Based Drought Prediction Layer

This layer receives the reduced data from the energy-saving fog layer and evaluates the intensity of drought for current and forthcoming situations. This layer is further composed of three layers: 1) drought severity assessment layer; 2) drought prediction layer; and 3) cloud storage layer, which are explained as follows.

*1) Drought Severity Assessment Layer:* Naïve Bayes (NB) classifier is used to assess the drought severity. This layer categorizes the drought severity as extreme drought (−4), severe drought (−3), moderate drought (−2), mild drought (−1), normal (0), mild wet (1), moderate wet (2), severe wet (3), and extreme wet (4). This classifier uses the Bayes theorem that works on the basis of conditional probability. Given the values of all the drought attributes for an event $E$, the values for the selected PCs are calculated as tuple $X$ in the dimension reduction layer. NB classifier presumes that the attributes are conditionally independent of each other. This condition is satisfied as the PCs are not correlated to each other. The tuple $X$ is considered to be a part of drought severity category $C_y$ only if $P(C_y/X) > P(C_z/X)$; $-4 \leq z \leq 4; z \neq y$, where $P(C_y/X)$ is probability that considered event $E$ is of drought severity category $C_y$ when chosen PCs have value $X$ and is calculated as

$$P\left(\frac{C_y}{X}\right) = \frac{P\left(\frac{X}{C_y}\right)P(C_y)}{P(X)} \tag{17}$$

where $P(X/C_y)$ is the probability that chosen PCs have value $X$ when event $E$ belongs to drought severity category $C_y$ and is determined as

$$P\left(\frac{X}{C_y}\right) = \sum_{m=1}^{M} P\left(\frac{x_m}{C_y}\right) = P\left(\frac{x_1}{C_y}\right) * P\left(\frac{x_2}{C_y}\right) * \ldots * P\left(\frac{x_M}{C_y}\right) \tag{18}$$

where $M$ is the total number of selected PCs and $x_m$ is the value of the $m$th PC. The whole procedure is explained in Algorithm 3. The results determined after the whole process are forwarded to the cloud storage layer.

**Algorithm 3** Drought Severity Category Determination

Input: Reduced Data set X of size Q*M having Q number of tuples and M number of chosen PCs.
Output: Drought Severity Category
Step 1. for y = −4 to 4
Step 1.1 Calculate Posterior Probability $P\left(\frac{C_y}{X}\right)$ using equations (17) and (18)
Step 2. Drought Severity Category of E = $\arg max_y( P\left(\frac{C_y}{X}\right))$
Step 3. Exit

**Algorithm 4** Drought Severity Prediction

Input: Previous Data for Drought Severity.
Step 1: Calculate the values of N, d and D to get stationary data.
Step 2: Examine Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) Plots for the determination of values for i, j, I and J to get structure of SARIMA model.
Step 3: Determine AR and MA coefficient for Seasonal and Non-Seasonal parts.
Step 4: Select model with minimum MAE, MSE and RMSE values as the best fit model.
Step 5: Predict Drought Severity Category for time t using equation (20).
Output: Drought Severity Category for time t.

*2) Drought Prediction Layer:* The drought prediction layer determines the severity of future drought events by evaluating the current and past severity values as calculated by the drought assessment layer. The seasonal autoregressive integrated moving average (SARIMA) is used for this purpose. The ARIMA model, which is a mixture of autoregressive (AR) and moving average (MA), is extended by, including seasonal component to it to form seasonal ARIMA. The AR and MA models predict the value of drought severity by taking linear combination of the previous drought severity values and previous drought severity prediction errors, respectively. ARIMA $(i, j)$, a combination of AR($i$) and MA($j$), is calculated as

$$\begin{aligned} D^t_{\text{Severity}} = \mu &+ \phi_1 D^{t-1}_{\text{Severity}} + \phi_2 D^{t-2}_{\text{Severity}} \\ &+ \cdots + \phi_i D^{t-i}_{\text{Severity}} - \theta_1 \ \ \varepsilon_{t-1} - \theta_2 \ \ \varepsilon_{t-2} \\ &- \cdots - \theta_j \ \ \varepsilon_{t-j} + \varepsilon_t \end{aligned} \tag{19}$$

where

| | |
|---|---|
| $i, j$ | number of AR and MA terms, respectively; |
| $\mu$ | constant term; |
| $D^{t-1}_{\text{Severity}} \ldots D^{t-i}_{\text{Severity}}$ | previous drought severity values; |
| $\phi$ | autoregression coefficient; |
| $\varepsilon_t$ | random error at time; |
| $\varepsilon_{t-1}, \varepsilon_{t-2} \ldots \varepsilon_{t-j}$ | previous drought severity prediction errors; |
| $\theta$ | MA coefficient. |

As drought conditions highly depend on the seasons, it is necessary to incorporate seasonal elements in the ARIMA model to form the $SARIMA(i,\ d,\ j)(I,\ D,\ J)N$ model, where $I$, $D$, $J$, and $i$, $d$, $n$ are the numbers of AR, differencing, and the MA terms for seasonal and nonseasonal component, respectively. $N$ is the time period of repeating seasonal pattern. The multiplicative SARIMA model is obtained as

$$\phi_{\mathrm{AR}}(B)\phi_{\mathrm{SAR}}\left(B^N\right)\left(1-B^N\right)^D D^t_{\mathrm{Severity}} = \theta_{\mathrm{MA}}\ \ (B)\ \theta_{\mathrm{SMA}}\ \ \left(B^N\right)\varepsilon_t \tag{20}$$

where

$B$ backshift operator;
$\phi_{\mathrm{SAR}}$, $\phi_{\mathrm{AR}}$ seasonal and nonseasonal AR coefficients;
$\theta_{\mathrm{SMA}}$, $\theta_{\mathrm{MA}}$ seasonal and nonseasonal MA coefficients.

The whole process for the prediction of the drought severity level ($D^t_{\mathrm{Severity}}$) comprises steps illustrated in Algorithm 4. Among all the estimated SARIMA models, the model with the minimum root-mean-square error (RMSE), mean square error (MSE), and mean absolute error (MAE) values is used to predict the drought severity value.

*3) Cloud Storage Layer:* This layer saves the results and evaluations of drought assessment for current and forthcoming events so that these can be utilized properly.

### D. Information Dissemination Layer

The agencies dealing with drought management can access the results stored at the cloud storage layer so that actions can be taken beforehand to tackle with the situation.

## IV. Experimental Evaluation

This section elaborates the experimental setup for the proposed framework and its evaluation. It consists of five parts: 1) data accumulation; 2) energy saving; 3) assessment using NBs classifier after dimension reduction using KPCA; 4) prediction using the SARIMA model; and 5) performance analysis.

### A. Data Accumulation

Data for various drought variables of three different climate divisions in Texas, USA, were retrieved from their official information centers. Temperature and precipitation data were taken from the official website of the National Centre for Environmental Information [28]. The data set for groundwater, streamflow, evapotranspiration, and drought was obtained from the website of the Texas Water Development Board [29], the United States Geological Survey [30], Texas A&M AgriLife Extension [31], and the United States Drought Monitor [32], respectively. In addition to this, data for soil moisture and soil temperature were taken from the Natural Resource Conservation Service [33]. All these data sets are stored and integrated at Amazon EC2 which provides infrastructure as a service (IaaS) [34]. An instance "m1.xlarge" of Amazon machine image (AMI) is run on CentOS 6.7 with a Linux 2.6.32 Xen Kernel.

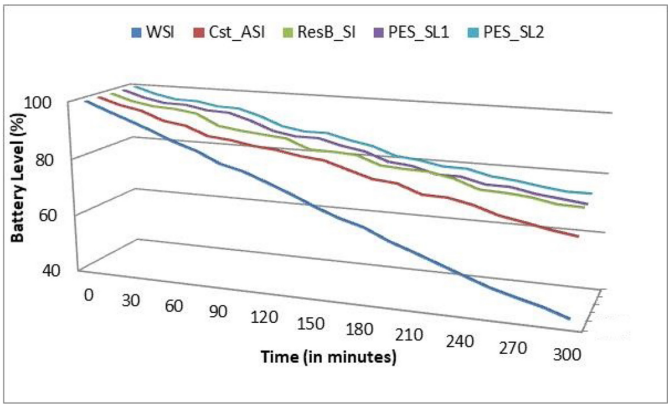


Fig. 3. Battery life of sensor node.

TABLE IV
Comparison of Accuracy and Battery Consumption of Energy Efficiency Schemes

| | WSI | Cst_ASI | ResB_SI | PES_SL1 | PES_SL2 |
|---|---|---|---|---|---|
| Battery Level% | 42.13 | 66.19 | 73.35 | 72.34 | 73.12 |
| Error Rate | 0 | 0.415 | 3.75 | 0.083 | 0.086 |
| CEC (nJ*$10^6$) | 277 | 146 | 130 | 134 | 133 |

### B. Energy Saving

Various aspects of the proposed energy-saving model are investigated by comparing it with three other schemes.

Case 1: Without sleep interval (WSI).
Case 2: With constant active and sleep interval (Cst_ASI).
Case 3: Sleep interval is calculated by taking the residual battery into account (ResB_SI).
Case 4: Proposed energy-saving system with SL = 0.05 (PES_SL1).
Case 5: Proposed energy-saving system with SL = 0.01 (PES_SL2).

In ResB_SI, sleep interval is increased as the battery level decreases. The proposed energy-saving method is examined for two different values of SL to investigate the effect of SL value on accuracy, energy consumption, and battery life of sensors. For the practical analysis, 15 humidity sensors were deployed in the university campus. All the sensor nodes send data to the android-based mobile phone which acts as an enode. The sensor nodes are connected to the Arduino board which sends the data to the enode using its ESP8266 Wi-Fi module. For each of the above cases, one enode having three humidity sensors executing the respective energy-saving scheme was taken. All the four enodes forward the collected data to the server placed at a university campus that works as the gateway and sends the data to Amazon EC2.

In the 300-min experiment, the residual battery level of the sensors in the data sensing layer under each case is analyzed. Fig. 3 shows that WSI has the lowest battery level with a value 42.13% as it is active all the time. WSI is followed by Cst_ASI with a battery level 66.19%. ResB_SI has the highest battery level of 73.35% as it considers only residual battery power to predict the next sleep interval while PES_SL1 has 72.34% of battery left. The accuracy and battery consumption of all the cases have been evaluated in Table IV. The most

TABLE V
ASSESSMENT ACCURACY EVALUATION

| | Accuracy | | | | Sensitivity | | | | Specificity | | | | Precision | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | NB | KNN | SVM | DT | NB | KNN | SVM | DT | NB | KNN | SVM | DT | NB | KNN | SVM | DT |
| KPCA-Guassian | 94.86 | 92.67 | 93.8 | 91.29 | 90.59 | 91.05 | 88.79 | 85.43 | 96.5 | 94.79 | 95.73 | 94.36 | 90.86 | 94.16 | 88.88 | 85 |
| KPCA-Polynomial | 94.79 | 91.86 | 93.2 | 91.14 | 90.12 | 90.8 | 88.61 | 84.31 | 96.4 | 93.56 | 95.6 | 94.16 | 90.77 | 94.11 | 88.81 | 84.85 |
| KPCA-Laplacian | 94.78 | 91.92 | 93.3 | 91.04 | 90.08 | 90.9 | 88.45 | 84.25 | 96.4 | 93.48 | 95.62 | 94.18 | 90.65 | 94.09 | 88.62 | 84.91 |
| PCA | 92.45 | 89.05 | 91.95 | 89.1 | 88.94 | 89.45 | 87.71 | 82.86 | 95.02 | 92.17 | 94.15 | 91.98 | 88.95 | 92.13 | 87.12 | 82.69 |

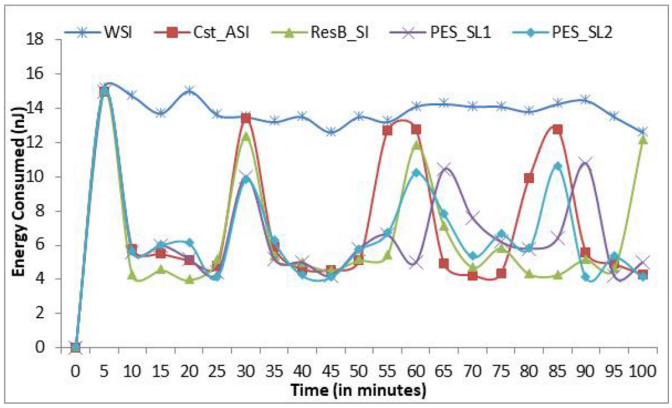


Fig. 4. Energy consumption of enode.

accurate system is WSI but it shows the lowest battery level. ResB_SI has the highest battery level but it is the least accurate system. PES_SL1 is a highly accurate system with a 0.083% error rate and it reveals the difference of just 0.99% in the battery level as compared to ResB_SI. The proposed method can meet the high accuracy requirements by compromising little on battery consumption. However, the battery consumption can be further reduced by decreasing the value of SL as a lower value of SL will result into higher threshold value and hence less active time. It also affects the accuracy with a small difference as it happens in PES_SL2. Furthermore, the energy consumption by enodes at the fog layer is examined in Fig. 4. It reveals that there is a sudden increase in energy consumption by enode initially because all the sensor nodes are configured in the beginning. In WSI, all the nodes are active and sending data continuously. Hence, the fluctuation is minimal. But the energy consumption changes a lot in the proposed method and the rest of the schemes as enode consumes more energy when the sensor nodes are in the active mode as it receives more amount of data. Cst_ASI has Cst_ASI, hence, the graph shows the constant gaps between the peaks of active periods. In ResB_SI, the gaps between the peaks increase with the time because the sleep interval increases with the time as the battery level decreases. Table VI shows the cumulative energy consumption (CEC) at enodes after 100 min in all the considered schemes. It reveals that the proposed method saves a significant amount of energy at the fog layer as compared to WSI and Cont_ASI. Overall the proposed energy-saving method saves power consumption at the data sensing layer as well as at enode in the fog layer with minor difference in accuracy.

### C. Assessment Using Naïve Bayes Classifier After Dimension Reduction Using KPCA

The dimensionality of the integrated data set received from the data collection layer is reduced using KPCA in *R* Studio by installing "kernlab" package from the CRAN repository. Gaussian radial basis function is used as the kernel function because of its performance in many applications [35], [36]. First, three PCs make 93.87% variance cumulatively and hence, are selected for further analysis. After KPCA, the NBs classifier is applied to the resultant data set in AMI instance where the drought severity category is specified. The combination of the KPCA Gaussian kernel function and the NB classifier is investigated for its performance by comparing it with combination of two other kernel functions of nonlinear KPCA, namely, KPCA polynomial, KPCA Laplacian, and linear PCA with classifiers, including decision tree (DT) [37], support vector machine (SVM) [38], [39], and $K$-nearest neighbor (K-NN). The performance is evaluated using metrics accuracy, specificity, precision, and sensitivity. Table V shows that the NB classifier performs better in determining the severity level of the drought with the highest accuracy and specificity with KPCA Gaussian kernel.

Furthermore, to show the effectiveness of the proposed method and its ability to capture the variability of factors, including temperature, groundwater, soil moisture, and precipitation, it is compared with extensively used drought indices, namely, SPI-1, SPI-6, SPI-12, and PDSI, by calculating the correlation coefficients as shown in Fig. 5. The comparison of drought indices shows that SPI-12 reveals a maximum correlation with groundwater and SPI-1 has a maximum correlation with the rest of the drought causing attributes. The proposed framework outdoes all drought indices with slightly greater correlation values. The correlation value with temperature, precipitation, soil moisture, and groundwater is the highest for the proposed model with values $-0.16$, 0.61, 0.67, and $-0.62$, respectively, in case of high plains climate division. Similarly, the proposed system performs better in low rolling plains and East Texas climate divisions with values $-0.14$, 0.65, 0.67, and $-0.53$ (low rolling plains) and $-0.13$, 0.63, 0.64, and $-0.61$ (East Texas) for temperature, precipitation, soil moisture, and groundwater, respectively. Therefore, the proposed system possessed the best performance by demonstrating substantial correlations. It reveals that the proposed system represents associations between the drought severity level and causing variables clearly.

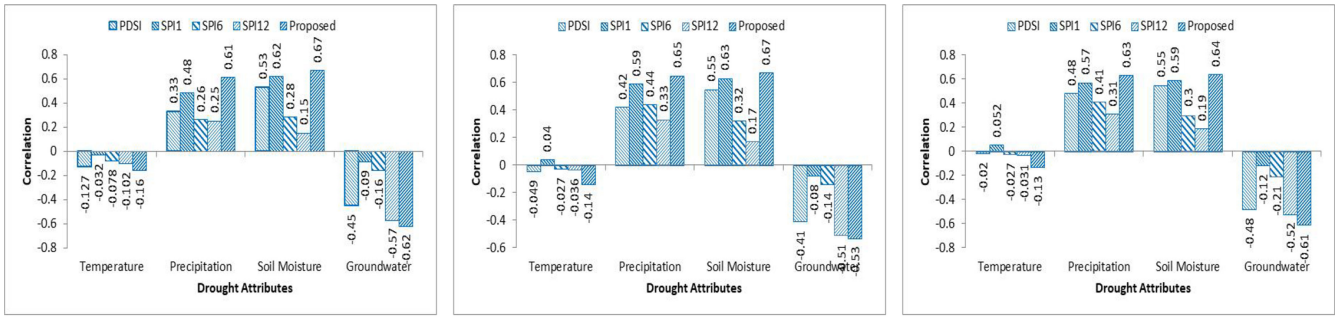


Fig. 5. Correlation of drought indices and proposed system with drought inducing attributes for (a) high plains, (b) low rolling plains, and (c) East Texas.

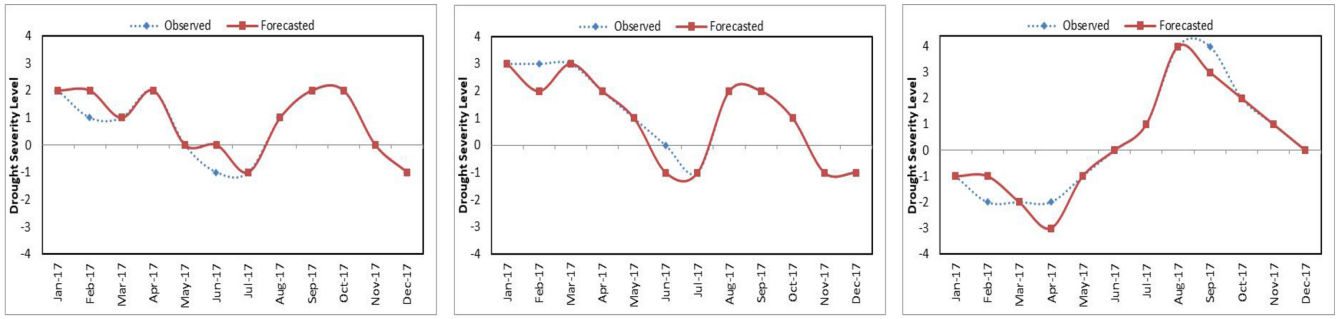


Fig. 6. Monthly prediction for (a) high plains, (b) low rolling plains, and (c) East Texas.

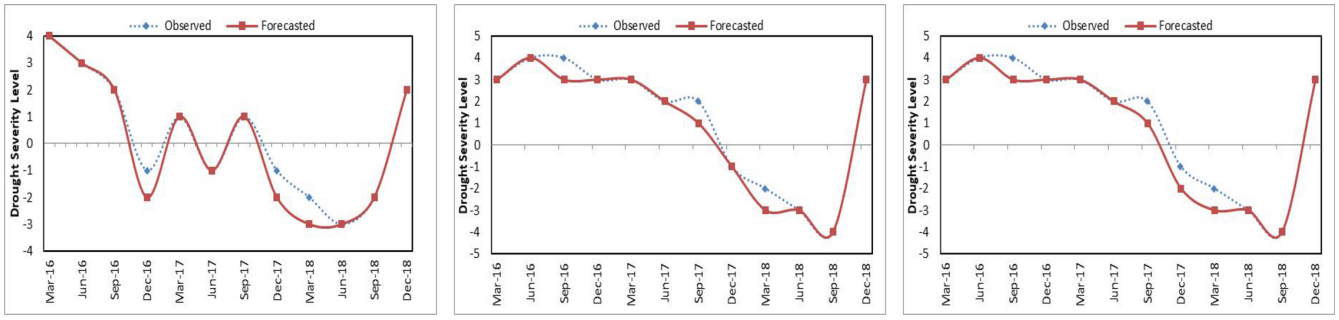


Fig. 7. Quarterly prediction for (a) high plains, (b) low rolling plains, and (c) East Texas.

TABLE VI
PREDICTION ACCURACY EVALUATION

| Time Frame | Climate Division | MAE | MSE | RMSE |
|---|---|---|---|---|
| Monthly | High Plains | 0.16667 | 0.16667 | 0.408248 |
| | Low Rolling Plains | 0.16667 | 0.16667 | 0.408248 |
| | East Texas | 0.25 | 0.25 | 0.5 |
| Quarterly | High Plains | 0.25 | 0.25 | 0.5 |
| | Low Rolling Plains | 0.333333 | 0.333333 | 0.57735 |
| | East Texas | 0.333333 | 0.333333 | 0.57735 |

### *D. Prediction Using SARIMA Model*

The results produced by NB classifier are given as an input to the SARIMA model which is implemented using "sarima"—an *R* package on *R* Studio present in Amazon EC2. "sarima" consists of a broad variety of functions for time-series modeling with ARIMA. In order to determine the most accurate structure of the SARIMA model and to test its performance spatially, ACF and PACF plots for monthly time-series data of three climate divisions of Texas are investigated. A set of suitable SARIMA models is determined for each climate division and the performance of each one of them is examined to select the final model. SARIMA (0,1,1)(0,1,1)12, SARIMA (0,1,2)(0,1,1)12, and SARIMA (1,1,0)(0,1,1)12 are the final models for high plains, low rolling plains, and East Texas climate divisions of Texas, respectively. Fig. 6 shows the monthly prediction of drought severity for all the three climate divisions.

Moreover, the performance of SARIMA model is examined for another time frame. The SARIMA models were determined by following the same process for quarterly prediction. SARIMA (0,1,1)(1,1,0)4 is the final model for high plains and

East Texas climate division while SARIMA (0,1,2)(1,1,0)4 is the final model for low rolling plains. Fig. 7 shows the quarterly drought severity prediction for all the considered climate divisions. Further more, the performance measures MAE, MSE, and RMSE are calculated for monthly and quarterly prediction in each climate division as shown in Table VI. This reveals that the SARIMA model has made this whole system highly accurate with very low error values but the accuracy falls as the prediction period increases from monthly to quarterly.

## V. Prerequisites for System Implementation

The sensors with good working condition should be deployed for the full area coverage to monitor drought conditions continuously. Reliable sensor-fog and fog-cloud communication along with the fog nodes having sufficient computing and storage capacity to receive and process data from connected sensor nodes are required. The initial values of sleep and active interval of sensors can vary according to the climatic season.

## VI. Conclusion

In this article, a fog-assisted architecture for drought assessment and prediction was proposed. This system has efficiently determined the accurate drought intensity category for the study area by considering a huge stretch of drought causing attributes. The main contribution of this framework is the optimization of the amount of data collected and hence, increasing the sensor lifetime by using methods based on data variance. Moreover, KPCA is employed at the fog layer for dimension reduction to utilize network bandwidth optimally. The proposed system can assess the drought severity level for future time periods of different scales efficiently. Moreover, it presents the performance evaluation of the proposed framework and other drought indices in indicating the changes of drought causing elements. The results are stored at cloud storage for the drought management agencies so that suitable approaches can be followed for timely action to alleviate the destruction.

## References


[1] G. Monacelli, M. C. Galluccio, and M. Abbafati, "Drought assessment and forecasting," in *Drought Within the Context of the Region VI*, 2005.

[2] P. Wallemarq, R. Below, and D. McLean, *UNISDR and CRED Report: Economic Losses, Poverty & Disasters (1998–2017)*, Centre Res. Epidemiol. Disasters, Brussels, Belgium, 2018. [Online]. Available: http://www.unisdr.org/files/61119_credeconomiclosses.pdf

[3] Y. Bayissa *et al.*, "Comparison of the performance of six drought indices in characterizing historical drought for the Upper Blue Nile Basin, Ethiopia," *Geosciences*, vol. 8, no. 3, p. 81, 2018.

[4] S. Fang *et al.*, "An integrated system for regional environmental monitoring and management based on Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1596–1605, May 2014.

[5] B. Kantarci and H. T. Mouftah, "Trustworthy sensing for public safety in cloud-centric Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 360–368, Aug. 2014.

[6] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.

[7] W. Shi *et al.*, "Effective prediction of missing data on Apache Spark over multivariable time series," *IEEE Trans. Big Data*, vol. 4, no. 4, pp. 473–486, Dec. 2018.

[8] G. Wang, J. Qiao, J. Bi, W. Li, and M. Zhou, "TL-GDBN: Growing deep belief network with transfer learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 874–885, Apr. 2018.

[9] S. Poornima and M. Pushpalatha, "Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network," *Soft Comput.*, vol. 23, no. 18, pp. 8399–8412, 2019.

[10] Y. W. Soh, C. H. Koo, Y. F. Huang, and K. F. Fung, "Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia," *Comput. Electron. Agricult.*, vol. 144, pp. 164–173, Jan. 2018.

[11] M. Ali, R. C. Deo, N. J. Downs, and T. Maraseni, "Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting," *Comput. Electron. Agricult.*, vol. 152, pp. 149–165, Sep. 2018.

[12] J. Yu, J. Lim, and K. S. Lee, "Investigation of drought-vulnerable regions in North Korea using remote sensing and cloud computing climate data," *Environ. Monitor. Assessment*, vol. 190, no. 3, p. 126, 2018.

[13] W. Zou, G. Li, and W. Yu, "MapReduce functions to remote sensing distributed data processing—Global vegetation drought monitoring as example," *Softw. Pract. Exp.*, vol. 48, no. 7, pp. 1352–1367, 2018.

[14] K. Du, Z. Sun, F. Zheng, J. Chu, and J. Ma, "Monitoring system for wheat meteorological disasters using wireless sensor networks," in *Proc. Annu. Int. Meeting Amer. Soc. Agricult. Biol. Eng. (ASABE)*, 2017, p. 18.

[15] G. B. Demisse *et al.*, "Information mining from heterogeneous data sources: A case study on drought prediction," *Information*, vol. 8, no. 3, p. 79, 2017.

[16] Z. Hao, F. Hao, V. P. Singh, W. Ouyang, and H. Cheng, "An integrated package for drought monitoring, prediction and analysis to aid drought modeling and assessment," *Environ. Model. Softw.*, vol. 91, pp. 199–209, May 2017.

[17] P. Maca and P. Pech, "Forecasting SPEI and SPI drought indices using the integrated artificial neural network," *Comput. Intell. Neurosci.*, vol. 2016, p. 14, Dec. 2016.

[18] R. Zhang, A. Nayak, and J. Yu, "Sleep scheduling in energy harvesting wireless body area network," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 95–101, Feb. 2019.

[19] G. Sakya and V. Sharma, "ADMC-MAC: Energy efficient adaptive MAC protocol for mission critical applications in WS," *Sustain. Comput. Informat. Syst.*, vol. 23, pp. 21–28, Sep. 2019.

[20] H. P. Gupta, S. V. Rao, and T. Venkatesh, "Sleep scheduling protocol for *k*-coverage of three-dimensional heterogeneous WSN," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8423–8431, Oct. 2016.

[21] H. Mostafaei, A. Montieri, V. Persico, and A. Pescap, "A sleep scheduling approach based on learning automata for WSN partial coverage," *J. Netw. Comput. Appl.*, vol. 80, pp. 67–78, Feb. 2017.

[22] W. Liu, Y. Shoji, and R. Shinkuma, "Logical correlation-based sleep scheduling for WSNs in ambient-assisted homes," *IEEE Sensors J.*, vol. 17, no. 10, pp. 3207–3218, May 2017.

[23] P. K. Sahoo, H. K. Thakkar, and I. S. Hwang, "Pre-scheduled and self organized sleep-scheduling algorithms for efficient *K*-coverage in wireless sensor network," *Sensors*, vol. 17, no. 12, 2017, Art. no. E2945.

[24] B. Singh and D. K. Lobiyal, "Traffic-aware density-based sleep scheduling and energy modeling for two dimensional Gaussian distributed wireless sensor network," *Wireless Pers. Commun.*, vol. 70, no. 4, pp. 1373–1396, 2013.

[25] B. Zebbane, M. Chenait, and N. Badache, "A group-based energy-saving algorithm for sleep/wake scheduling and topology control in wireless sensor network," *Wireless Pers. Commun.*, vol. 84, no. 2, pp. 959–983, 2015.

[26] M. O. Oladimeji, M. Turkey, and S. Dudley, "HACH: Heuristic algorithm for clustering hierarchy protocol in wireless sensor network," *Appl. Soft Comput.*, vol. 55, pp. 452–461, Jun. 2017.

[27] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor network," *Appl. Soft Comput.*, vol. 13, no. 4, pp. 1741–1749, Apr. 2013.

[28] NOAA's National Climatic Data Center (NCDC). *Divisional Time Series Data*. Accessed: May 31, 2019. [Online]. Available: https://www.ncdc.noaa.gov/cag/divisional/time-series
[29] Texas Water Development Board. *TWDB Groundwater Database*. Accessed: May 31, 2019. [Online]. Available: https://waterdatafortexas.org/groundwater/download
[30] United States Geological Survey. *Average StreamFlow Database*. Accessed: May 31, 2019. [Online]. Available: https://waterwatch.usgs.gov/index.php?id=ww
[31] Texas A&M Agrilife Extension. *Potential Evapotranspiratio*. Accessed: May 31, 2019. [Online]. Available: https://etweather.tamu.edu/pet/
[32] National Drought Mitigation Center. *United States Drought Monitor*. Accessed: May 31, 2019. [Online]. Available: https://droughtmonitor.unl.edu/Data/DataDownload.aspx
[33] Natural Resource Conservation Service. *Daily SCAN Standard Report—Period of Record*. Accessed: May 31, 2019. [Online]. Available: https://wcc.sc.egov.usda.gov/nwcc/rgrpt?report=daily_scan_por
[34] M. H. Ghahramani, M. Zhou, and C. T. Hon, "Toward cloud computing QoS architecture: Analysis of cloud systems and cloud service," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 1, pp. 6–18, Jan. 2017.
[35] Z. Xu *et al.*, "Software defect prediction based on kernel PCA and weighted extreme learning machine," *Inf. Softw. Technol.*, vol. 106, pp. 182–200, Feb. 2019.
[36] J. Li, S. Chu, and J.-S. Pan, "Kernel principal component analysis (KPCA)-based face recognition," in *Kernel Learning Algorithms for Face Recognition*. New York, NY, USA: Springer, 2014, pp. 71–99.
[37] S. Teng, N. Wu, H. Zhu, L. Teng, and W. Zhang, "SVM-DT-based adaptive and collaborative intrusion detection," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 1, pp. 108–118, Jan. 2018.
[38] P. Zhang, S. Shu, and M. Zhou, "An online fault detection model and strategies based on SVM-grid in cloud," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 445–456, Mar. 2018.
[39] W. Zhang, H. Zhang, J. Liu, K. Li, D. Yang, and H. Tian, "Weather prediction with multiclass support vector machines in the fault detection of photovoltaic system," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 3, pp. 520–525, Jul. 2017.