




Hybrid Ensemble Based Machine Learning for Smart Building Fire Detection Using Multi Modal Sensor Data

Sandip Jana and Saikat Kumar Shome ^{*}, CSIR - Central Mechanical Engineering Research Institute (CSIR-CMERI) Campus, Durgapur 713209, India and Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

Received: 23 April 2021/**Accepted:** 15 November 2022/**Published online:** 8 December 2022

Abstract. Fire disasters are one the most challenging accidents that can take place in any urban buildings like houses, offices, hospitals, colleges and industries. These accidents which the world faces now, have never been more frequent and fatal, leading to innumerable loses, damage of expensive equipment and unparalleled human lives. The concrete landscapes are threatened by fire disasters, which have prolifically outnumbered in the last decade, both in intensity and frequency. Thus, to minimize the impact of fire disasters, adoption of well planned, intelligent and robust fire detection technology harnessing the niches of machine learning is necessary for early warning and coordinated prevention and response approach. In this research a novel hybrid ensemble technology based machine algorithm using maximum averaging voting classifier has been designed for fire detection in buildings. The proposed model uses feature engineering pre-processing techniques followed by a synergistic integration of four classifiers namely, logistic regression, support vector machine (SVM), Decision tree and Naive Bayes classifier to yield better prediction and improved robustness. A database from NIST has been chosen to validate the research under different fire scenarios. Results indicate an improved classification accuracy of the proposed ensemble technique as compared to reported literatures. After validating the algorithm, the firmware has been implemented on a laboratory developed prototype of smart multi sensor, embedded fire detection node. The designed smart hardware is successfully able to transmit the sensed data wirelessly onto the cloud platform for further data analytics in real time with high precision and reduced root mean square error (MAE).

Keywords: Regression, Classifier, Model, Fire, Ensemble, Machine learning, Prediction

^{*}Correspondence should be addressed to: Saikat Kumar Shome,
E-mail: saikatksHOME@cmERI.res.in



1. Introduction

Fire knows no discrimination and is a paramount reason of crucial and catastrophic disasters around the globe which can occur in different environments like offices, industries, residential complex, schools, etc. Automated fire detection system offers the flexibility to assess essential physical and environmental parameters and their impact detection and prediction of fire either at an early stage or even prior to the outbreak. Accordingly, automatic fire detection systems have attracted considerable attention owing to its importance in reducing fire damage.

The best way to deal with disasters is to nip it off in the bud. To reduce the casualties and consequences of fire and minimize associated financial downside, prevention of the spread of fire is essential at a nascent stage. Fire detection plays a pivotal role in this parlance, to trigger timely warning signal reporting initiation of fire event.

Most of traditional building fire detection systems use off the shelf single sensor based fire detection with no intelligence whatsoever. This brings up two prominent bottlenecks which calls for research intervention—(a) reliable and accurate detection of fire occurrence, and (b) early prediction and warning system to forecast the occurrence of fire based on similar pre-conditions. An accurate and timely detection of fire is essential to mitigate the onset of false positive alarms raised by the fire detection system. The detection sensors must be able to differentiate and discriminate actual fire smoke from non-fire incidents as an inappropriate triggering of fire alarm not only causes disruptions in the production pipeline but also raises panic. At the same time, an early warning system based on AI which can predict the occurrence of fire shall facilitate pre-emptive scheduling of necessary activities, thus ensuring no fire linked damages. Therefore, in order to overcome the drawbacks associated with present fire alarm systems, it is necessary to develop and implement reliable and effective fire management systems to combat this disaster.

Interventions of technology in mitigating fire outbreaks have been studied extensively. Some researchers have worked with computer vision to analyse the fire images [1, 2] however it calls for expensive fire grade and far vision camera that entails considerable hardware cost. The advances of deep learning based technologies has played a prominent role in enhancing the quality of our lives in the past decade, as elaborated later, with a few contributions in fire technology as well. However, on the whole, AI and ML has a lot to offer as a promising technological solution to fire disaster management landscape. In a pioneering work, Brian et al [3] employs neural network approach for detecting and analyzing the fire signals and addressed false alarm conditions. At the same time, machine learning based classifications have been used for detection of fire problems in [4].

Soft computing techniques comes with the advantage of eliminating expensive hardware like image acquisition devices and have played a major role in fire outbreak mitigation [5–7]. Soft computing have also been successfully employed to bring out the underlying relationship between the causal variables and context sensitive fire occurrences [8–10] besides estimating the fire effected area linking future environmental conditions [11–13]. Quite a few researches have also been

implemented targeting prediction of fire using soft computation and learning technologies [14–20]. Artificial neural network (ANN) has also been adopted for early fire detection. Harnessing the potential of ANN and logistic regression, Bisquert et al. [21] and ref [22] reports a good classification accuracy achieved through this technique. Maeda et al. [23] identifies areas of high risk of fire incidents in Brazil employing ANN and the approach is suggestive of efficient detection. Several trade-offs are also required to be kept into consideration during implementing ANN algorithm such as nodes of hidden layer and number of nodes in each hidden layer [24]. Large number of training iterations may essentially over train the network, thus, negatively effecting the prediction accuracy [25]. Support vector machine (SVM) is another vertical of machine learning which have been widely used for fire detection and reported to have achieved good results and effective prediction capabilities [26–29]. An advantage of using SVM is that it does not require prior determination of probabilities, thus making it more preferable.

Machine learning based soft computation algorithms have also been used extensively in the recent past because of its efficient prediction capabilities. Few notable examples of the same as usage of learning in the form of random forest classifier [30–32], decision tree classifier [33, 34], support vector machine [35, 36], logistic regression [37], artificial neural network [38], Naïve bayes classifier [39] for classification and prediction tasks. Amongst all classifier approaches, ensemble based classifier has been seen to be more efficient [40] than any individual classifier as it learns from different aspects of training data considering features from the entire solution space [41]. Ref [42] presents a hybrid ensemble method for improved prediction of slope stability using ensemble classifiers and individual classifier technique. Weighted majority voting technique is used to combine the model and tenfold cross validation is used to validate the data for the slope prediction analysis in [43]. A weight based ensemble method WhmBoost is proposed in [44] for classifying balanced data in a binary classification task. The presented work uses two sampling methods and base classifiers with each of them being associated with the weight factor which results in better complementary advantages.

To reduce data imbalance, changing the learning process and modifying sensitivity of the algorithm, a hybrid method of data level approaches is implemented [45]. Hybrid ensemble methods [46–50] are more pronounced in favour of the minority class as it can separate the majority dataset from minority dataset in an effective way. Various sample technique can be adopted to improve the classification performance. The method of combining the sample technique and ensemble technique which leads to achieve desired performance in classification tasks primarily include adaboost [51, 52], voting [53], gradient boosting [54] approaches. Ref [55] represented a novel ensemble learning method which can detect forest fire in different scenarios. In this paper two individual classifiers Yolov5 and EfficientDet are used to detect the fire and another learner EfficientNet is used to reduce the false positive rate by 51.3% and an experiment is carried out on the dataset which can be signified that proposed ensemble learning method improves the detection performance by 2.5% to 10.9%. In [56], an ensemble model is developed which can produce exact solution and improves the feature selection than

multiple individual model. An experiment is conducted on hybrid MultiBoostAB Ensemble technique which has different feature selection for finding the model accuracy. The ensemble learning comprising with multiple learning algorithms is used to enhance the predictive performance of any model and hybrid ensemble learning method is a combination of multiple individual classifiers to solve a particular computational intelligence problem. In literature, several classification problems are investigated by using hybrid ensemble technique like classification in imbalance data [57], pulsar candidate classification [58], classification in medical databases [59] and multiclass classification problem of oilseed disease dataset [60]. Ref [61] improved the prediction of slope stability by using hybrid ensemble technique. D Rosadi et al. proposed a prediction of forest fire by using adaptive boosting ensemble classification method [62]. In this method decision tree and SVM individual classifier method are used and consider the public dataset to configure the model. An extreme gradient boost hybrid ensemble learning method is developed by Ying Xie et al. to predict the burn area of the forest fire using forest fire dataset [63]. Proposed ensemble technique for detection of burned area for forest fire has better than other individual classifier in term of prediction accuracy for large-scale fire occurrences. Therefore in literatures, hybrid ensemble learning methods are used in different classification problem and prediction of forest fire system. However for building fire detection cases are not deployed.

In this research, a novel machine learning based algorithm is proposed and validated on robust multi sensor data. The contribution of this work is design of a real time hybrid ensemble classifier which synergistically integrates four individual classifiers namely logistic regression classifier, support vector machine (SVM), Decision tree classifier and Naive Bayes classifier. After necessary pre-processing, the dataset is used in the study. An average voting ensemble technique is used for better prediction and seen to improve robustness of the learning algorithm. Ten-fold cross validation technique is chosen to compare the performance of the proposed machine learning algorithm under different fire scenarios. Results has been quantified using model accuracy, model precision, recall, receiver operating characteristic (ROC), area under curve (AUC), cumulative and individual importance of the parameter and error calculation. After validation of the proposed methodology, experiment has been carried out in the laboratory test bench setup using developed smart IoT sensor node prototype.

The paper is organised as follows: Sect. 1 System description and architecture of hybrid ensemble learning technique, Sect. 2 brief introduction of individual classifiers and proposed novel hybrid ensemble by average voting technique, Sect. 3: presents the research methodology followed by proposed machine learning algorithm, data collection, cross validation and the experimental set up, Sect. 4: results and discussion has been presented in Sect. 4.

2. System Description

2.1. Hybrid and Individual Ensemble Learning

Compared to the single model learner with only one hypothesis over the data, ensemble learning can consider multiple hypotheses, as seen in Figure 1.

Ensemble learning method is a class of machine learning which trains itself from multiple learning frameworks such as random forest, decision tree or other learning algorithm and combines them to get a new better learner. The multiple learner or base learners which are same models but get trained with different data/parameters by selecting best single learner. The final results of the ensemble technique can be illustrated by using voting, averaging or adaboost method, shown in Figure 1

Prediction capability of the combined model gives a better result compared to single model prediction. Ensemble model can be classified as Homogenous Ensemble Method and Heterogeneous ensemble method. Homogenous Ensemble Methods is constructed by multiple classifiers such as boosting, bagging and random forest etc. using different training dataset while Heterogeneous ensemble Methods is developed by different kind of learning algorithms, such as voting, stacking etc. and utilise the training dataset to develop multiple model.

In this present study, a hybrid ensemble technique is proposed for better prediction and enhanced accuracy. The hybrid ensemble method integration of four individual classifier algorithm comprising with based learners such as Logistic Regression, Decision Tree Model, Support Vector Machine, Naive Bayes model using average voting methodology. To begin with, individual classifier models have been trained for prediction. Then four machine learning models have been trained through a hybrid ensemble technique. The classification accuracies have been compared using the confusion matrices of each of the models. For validation, tenfold cross-validation has been carried out and accuracies of each of the four models have been observed. Performance of the optimum hybrid ensemble classifier has also been compared with each single classifier model which shows better accuracy in prediction and lower RMSE error than other classifier models.

2.2. Individual Classifier Model

In this research paper a hybrid ensemble model has been developed by using different classifier model (logistic regression, support vector machine (SVM), Decision tree and Naive Bayes classifier). The experimental data have been collected from sensor node which is then fitted into the laboratory test bench set up as shown in Figure 4 in the manuscript. After the collection of sensor data then data are preprocessed and a dataset has been prepared for the model configuration. The model is trained with 80% of the dataset and remaining data's are kept for testing purpose. After splitting the dataset, the model is fitted or trained to produce the outcomes. A tenfold cross validation technique has been introduced to increase the effectiveness of the model, therefore the training dataset is divided into 10 subsets from where 9 subsets are used for training and remaining one is used for predicting purpose. After that ensemble approach is used to develop

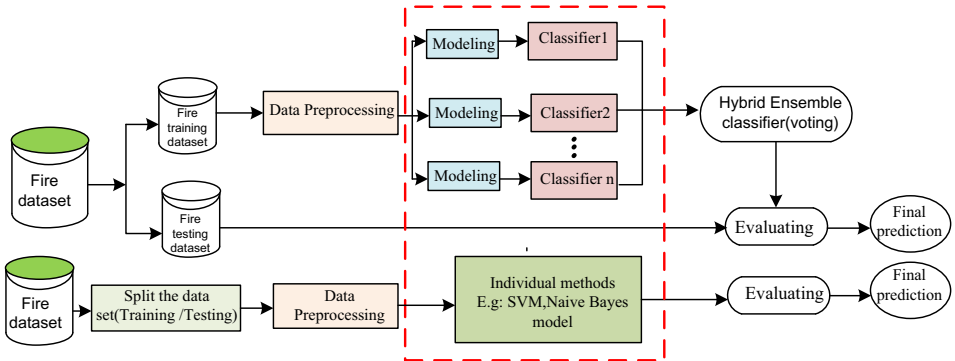


Figure 1. Comparison between individual and hybrid learning technique for fire detection.

more accurate ensemble classifiers model by addition of multiple number of individual classifiers.

Six individual classifier models have been trained with the real time sensors data collected through multi sensor node from experimental set up. Brief description of each of the classifier models is discussed below:

2.2.1. Logistic Regression Logistic regression is one of the machine learning algorithms which utilise the logistic function or sigmoid function and used for multi-class classification problems as well as binary classification problem. Logistic regression is a linear classifier therefore logistic function is defined as

$$f(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots\dots\dots\beta_r x_r \tag{1}$$

where, $f(x)$ dependent variable, x_1, x_2, \dots, x_r are explanatory variable and the variables $\beta_0, \beta_1, \dots, \beta_r$ are the estimators of the regression co-efficient or predicted weight.

2.2.2. Support Vector Machine Support vector machine classification algorithm is one of the most robust classification and regression algorithm, often used in several fields of application in science and engineering field. SVM plays an important role in the field of application of voice recognition, pattern recognition and also text categorisation. The main objective of support vector machine algorithm in binary classification is to get the minimum hyper planes which have maximum distance from the training data set. In nonlinear application, kernel function has been used to find the hyper plane which is represented by the non-linear decision boundary in the input spaces.

2.2.3. Decision Tree Classifier Decision tree algorithm is a machine learning technique which is used to find the data in replacement statistical procedures and to extract the decision. Different kinds of decision algorithm have been used to obtain their accuracy and cost effectiveness. A decision tree is a flow chart like

tree structure which includes branches, root node and leaf node. Internal node represents feature or attribute of the classifier, branches represents outcome or decision rule of a test and each leaf node denotes a class label. The top most of the tree is referred to as root node of decision tree, as seen in Figure 2.

2.2.4. *Naive Bayes Model* Naïve bayes classification is basically is used multi-label learning problem. A naive bayes classifier is related with the bayesian network, as shown in Equation 2 where C denotes single class variable and ‘n’ represents attributes of variables of X_I . Therefore ‘c’ is a class label variable and x_i represents a value of an attribute X_i . A naïve Bayes distribution can be represented as

$$P_r(c, x_1, \dots, x_n) = P_r(c) \prod_{i=1}^n P_r(x_i|c) \tag{2}$$

where, $P_r(c)$ and $P_r(x_i|c)$ are represented as class prior and conditional distribution.

2.3. Hybrid Ensemble Classifier

In the literature most of the proposed ensemble methods are developed by a single base estimator or single sampling method but with mixing the number of base estimator and number of sampling method which can give the system better performance.

The main objective of hybrid ensemble approach is to develop more accurate ensemble classifiers by addition of multiple numbers of individual classifiers. Ensemble classifier method combines the prediction of several base estimators to improve robustness of the system over the individual estimator. However, it is not certain that hybrid ensemble classifier shall always perform better than individual classifier however accuracy of hybrid ensemble classifier is always better than average accuracy of all single classifiers. There are many methods available in literature to develop hybrid ensemble classifier. The most widely used and computationally inexpensive method is majority voting and average voting.

In this research paper, average voting method has been implemented in real time to build several base estimators independently, after considering their averaged prediction. It is seen that performance of the combined estimator is better than of any of the single base estimators. A general architecture of the hybrid average voting classifier is shown in Figure 3 where the input dataset are pre-processed and followed by the intermediate base estimator and combining the logistic regression, support vector machine, decision tree classifier and naive bayes classifier models using average voting technique. All of the combined classifiers follow the probability rule of the average voting techniques. In this technique all individual classifier creates its on hypothesis (H1, H2, H3, H4) accordingly and for every output class a probability has been generated after that a best probability class has been selected for the final prediction due to the hybrid ensemble technique shown in Figure 3.

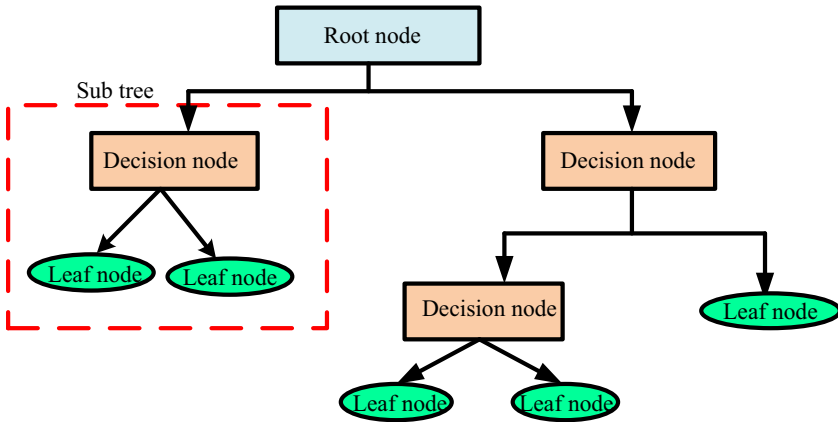


Figure 2. Basic algorithm structure of decision tree classifier.

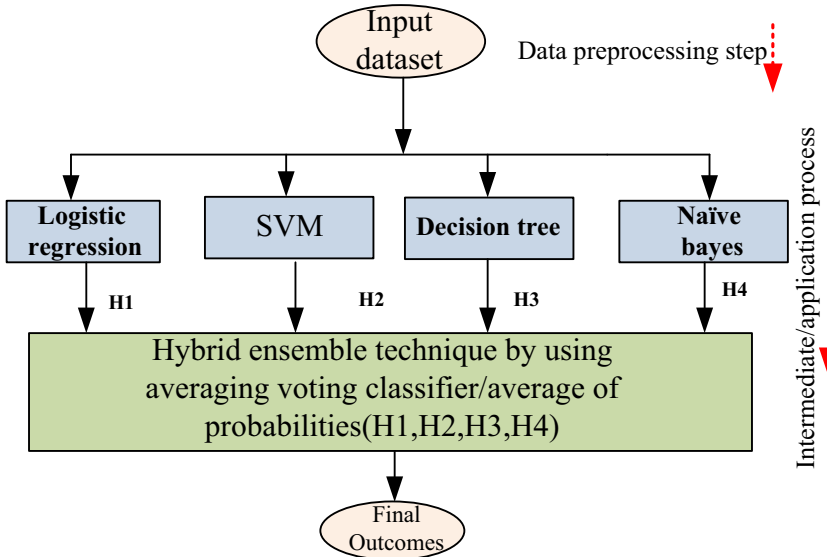


Figure 3. Proposed hybrid ensemble classifier model.

3. Research Methodology

In this section, a description of individual classifier and hybrid classifier ensemble techniques are used for fire prediction. The research methodology consists of three parts: Dataset preparation, novel machine learning algorithm design and cross validation of dataset.

3.1. Proposed Machine Learning Algorithm for Fire Detection

In recent research trends a hybrid ensemble learning techniques enhances more interested in the field of predictive modelling and it is combined the various learning classifier so that it improves the prediction accuracy over the single classifier model [64]. In this research a voting technique is used that combines the results of the multiple classifier model and weight are determined by gating network and the input of the model which has been created and base model are same and returns a weight to each of the base model in [65]. Two voting technique are mainly used like majority voting and average voting. In majority voting technique 50% vote are consider for final prediction and in average voting, the vote of the individual classifier has been averaged then predict the final decision. In this work we are considering average voting for combining the classifier and a general architecture of the hybrid average voting classifier shown in Figure 4.

A correlation coefficient denotes the strong relationship between two input variables. There are different kinds of correlation coefficients but here Pearson's coefficient has been used denoted by ρ due to its advantages.

Pearson's coefficient is defined as covariance between two input variables divided by the product of standard deviation

$$\rho(X, Y) = \frac{COV(X, Y)}{\sigma_X \sigma_Y} \quad (3)$$

$$\rho(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (4)$$

where, μ_X, μ_Y are mean of X and mean of Y.

A co-relation matrix has been obtained to visualize the relationship between sensor input data and labelled output data. Figure 5 indicates that the variable of dataset is distributed and the distribution of variable is not symmetric in nature. Variable range normally lies between [0 1] on their minimum and maximum values to improve the computation efficiency of the classifier. The correlation variable ranges are varied from -1 to 1 which corresponds to maximum positive correlation to maximum negative correlation. In this work, maximum and minimum range of sensor input of co-relation variable is -0.85 to 0.36 . The CO_2 and O_2 of the sensor data output variable has strongly correlated each other is shown in Figure 5.

3.2. Cross Validation and Performance Measures

K fold cross validation technique has been used for the prediction system to reduce the bias resulting from the random selection of training data and hold out data samples which has been used in [66]. In this paper, tenfold cross validation has been introduced, therefore the training dataset is divided into 10 subset from where 9 subsets are used for training and remaining one is used for predicting purpose. The training and prediction process has been iterated for 10 times with

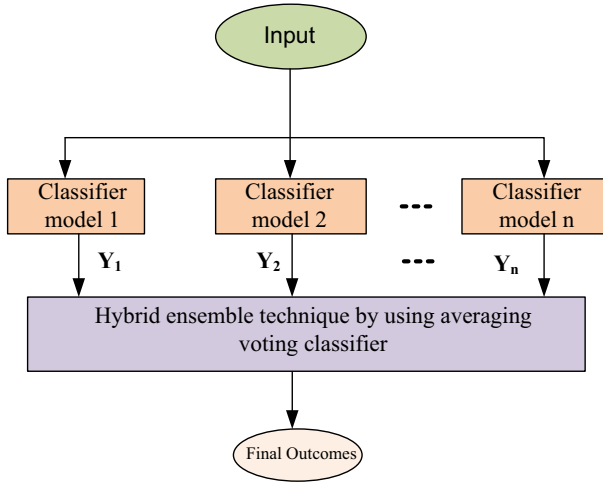


Figure 4. General architecture of the hybrid average voting classifier.

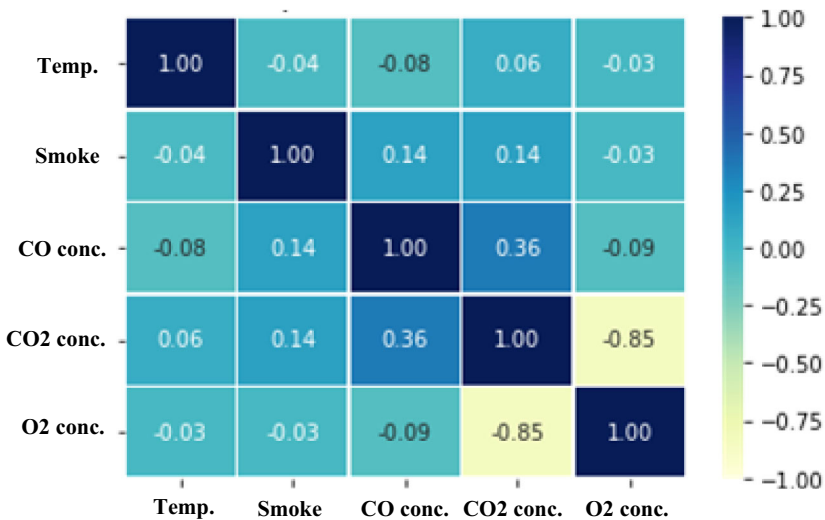


Figure 5. Correlation matrix of input and output variable of fire detection in dataset.

different subsets used as the predicting set. Finally, the performance of the prediction has been investigated by averaging the performance of training and predicting dataset. Performance has been measured by calculating the model accuracy, ROC curve and AUC. The performance of prediction can be portrayed by the confusion matrix shown in Figure 10 and the tenfold cross validation shown in Figure 6.

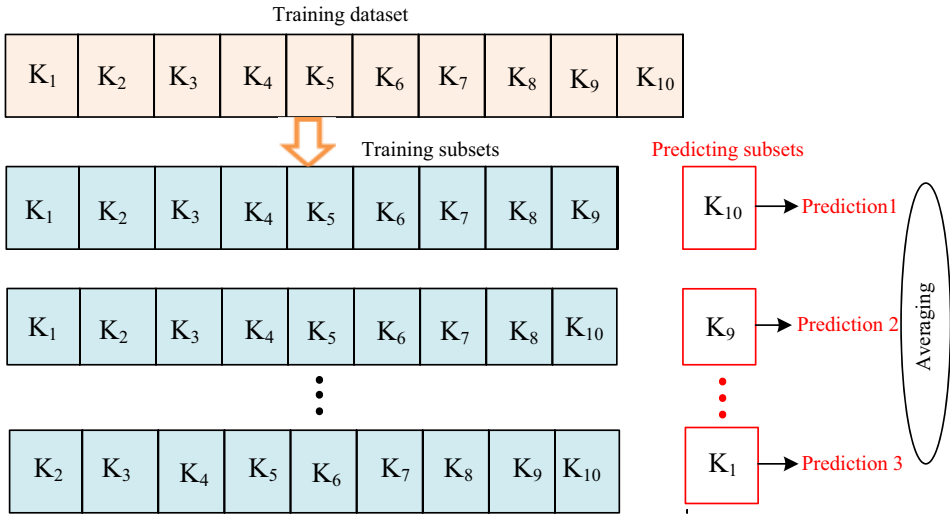


Figure 6. Tenfold cross validation of the obtained dataset.

3.3. Dataset Preparation

The fire data from the NIST Website “<https://www.nist.gov/el/nist-report-test-fr-4016>” has been considered for performance evaluation of the proposed model. On this dataset, we have applied the proposed hybrid ensemble based machine learning for validation it using five fire scenarios), two for smoldering fire dataset (SDC1, SDC3), two for flaming fire dataset (SDC5, SDC15) and one for cooking oil fire dataset (SDC12) conducted in a mock-up of a small house or apartment. At multiple positions within the data structure, concentrations of CO, CO₂, and O₂ were measured, as well as smoke and temperature. Details of the dataset are available in the referred website.

In Table 1, Precision, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of each machine learning model under different fire scenarios have also been investigated. Performance of the proposed work with similar reported research for fire detection system has been illustrated in Table 4. The results indicate an improvement in performance of the proposed model along with considerable performance in different fire scenarios. In the dataset, fire label column has labelled with “0” represented as non fire case and “1” labelled for fire case condition under different test cases.

Experimental validation of the proposed algorithm has also been carried out using the developed sensor node in a laboratory prototype as shown in Figure 8. In this paper, different gas sensors (MQ 3, MQ135, MQ-2) are used in the sensor node of the experimental setup to detect the fire. MQ 3 gas sensor is highly sensitivity to alcohol and MQ135 gas sensor has been used to detect NH₃, NO_x, alcohol, Benzene, smoke, CO₂, etc. MQ-2 gas sensor is a semiconductor sensor for combustible gas has high sensitivity to H₂, LPG, Propane and CO gas. The exper-

Table 1
Proposed Hybrid Machine Learning Models Under Different Fire Scenarios (individual dataset) and a Mixed Fire Scenario (merge dataset) in Term of Precision, MAE and RMSE

Fire scenarios	Precision	MAE	RMSE
Fire scenario 1(SDC1)	0.9545	0.1362	0.1362
Fire scenario2 (SDC3)	0.95432	0.1352	0.1352
Fire scenario3(SDC5)	0.94560	0.1256	0.1256
Fire scenario4(SDC12)	0.9632	0.1026	0.1026
Fire scenario(SDC15)	0.955	0.1310	0.1310
Fire scenario with merge dataset	0.9752	0.01009	0.10045

imental fire data such as temperature and gas concentration profile has been introduced as shown in Figure 7.

The smart fire sensor node comprises:

- *Embedded Controller board* Microchip ATmega328P Single board microcontroller (16 MHz Clock Speed with 32 KB in-system programmable flash) with Cloud Connectivity Chip.
- *Sensors* Different gas sensors (MQ135, MQ 2, MQ3) are used pertaining to fire outbreak eg smoke, CO₂, CO, O₂, etc. Temperature and Humidity sensors are also incorporated as it provides related pre-cursor information and the performance of gas sensor value may be improved by adjusting the load resistance value of the sensor.
- *Temperature and Humidity Sensor* DHT11 is an embedded humidity and temperature sensor provides signals in digital, I2C format useful in providing fire related pre-cursor information. Temperature and Humidity sensors are also incorporated as it provides related pre-cursor information.
- *Buzzer* As an actionable downlink based indicator of presence of fire event. The functionality of the buzzer is proposed to be extended and interfaced with a relay as an actionable counter-measure like switching-on of pump, etc. when presence of fire is affirmed by the cloud based analytics engine through downlink.

A laboratory scale test bed setup has been fabricated for experimentation of different fire conditions. The test bed is primarily automated in nature and comprises two chambers—one for electrical fire (right side), and other for gas-linked fires (left side). The chamber on the right side resembles common electrical fire and is powered on through a control switch which turns on and ignites an electric coil through a step down transformer. An electrical blower is also attached to bring down the flame. The left chamber is designed to experiment fires occurring through presence of several inflammable gasses. Sensing such conditions both pre and post fire scenarios reflects presence of crucial gasses and physical environmental factors which are indispensable in lending valuable insights for effective fire management. The setup also has means of extinguishing the fire through piped

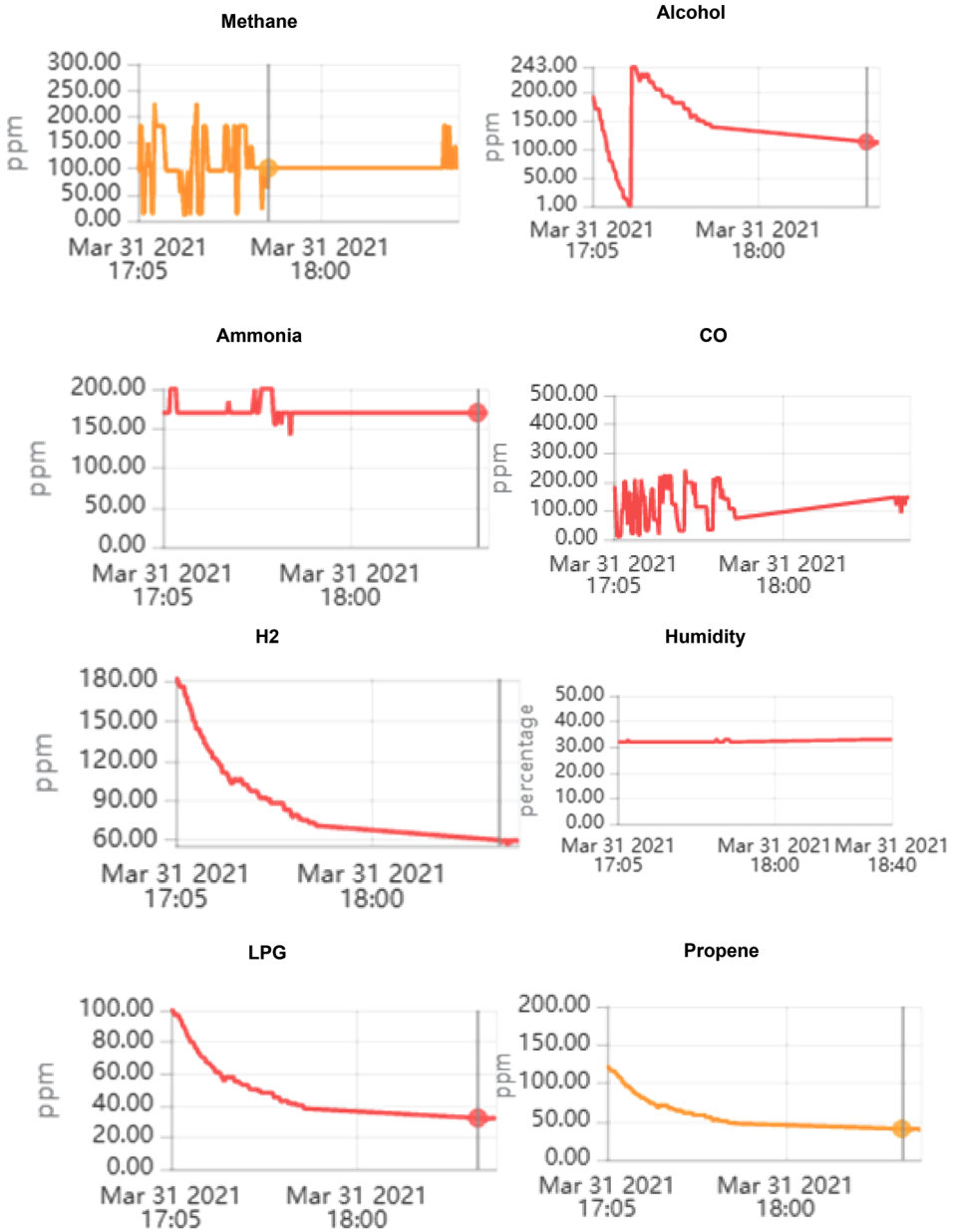


Figure 7. Experimental gas concentration result from gas sensor.

CO₂ release after the experiment is over. The developed smart and wireless multi sensor nodes can be placed in any of the chambers (in left chamber in Figure 8) and automatically detects the presence of flame, gas, and fire conditions and shall transmit to the base station for onward transmission to the cloud.

3.4. Internet of Things (IoT) Framework Used in the Study

Real time fire detection framework is essential to ultimately save life and prevent catastrophic disasters. After sensing physical parameters related to fire event, IoT system is essential not only for proper detection of the fire using cloud computation based advanced machine learning algorithms, but also to take preemptive and timely counter measure to mitigate the disaster. After validating the proposed algorithm on the NIST dataset, a small experiment has been carried out using lab level test bench set up using developed smart sensor node. Real time data from the wireless smart nodes are sent to the cloud platform for further data analytics using IoT chain uplink as well as to send automated alarms to beneficiaries. This entire system flow is a part of proposed smart fire detection setup proposed in this research. The results establish an end-to-end working prototype of an intelligent and smart fire detection framework using IoT chain. An IoT enabled architecture is implemented for real time management of fire situation, comprising four major components—sensors, networking, cloud and application server, and respective layers of communication protocols are shown in Figure 9.

- (a) *Sensors* They pertain to devices which detect the presence of certain fire related physical elements in the environment. Multiple parameters pertaining to fire outbreak can be captured through such sensors. Fire and smoke sensors provides valuable insights on the intensity of fire. The gas sensors like CO, CO₂ and O₂ help assess and develop the intelligent AI based framework by lending valuable information about pre-conditions and fire associated parameters which may be useful in forecasting an outbreak.
- (b) *Networking* There are several networking and communication technologies which can be used for transmission of acquired sensor data to the cloud. The commonly used ones are cellular (2G, 3G, 4G, etc.), radio frequency (LoRa, Zigbee, etc.), or inexpensive WiFi. These technologies vary in terms of offered performance like transmission range, data latency, power consumption, battery shelf life, etc. and their implementation depends on the specific requirement keeping into consideration, local conditions. The primary components of networking are data loggers, repeaters or gateways depending on the coverage of the local network area and required coverage.
- (c) *Cloud* Data received from the sensors through the internet needs to be stored on cloud framework for future usage. The cloud platform either public or private can host multiple applications, enabling sensor device management, configuration, and routing.
- (d) *Application Server* This is the last stage of the IoT chain and focuses on advanced analytics suitable for the fire management application. Data visualizations along with customised dash-boards offer unprecedented insights through diverse use-cases facilitating predictive management and fire projections.

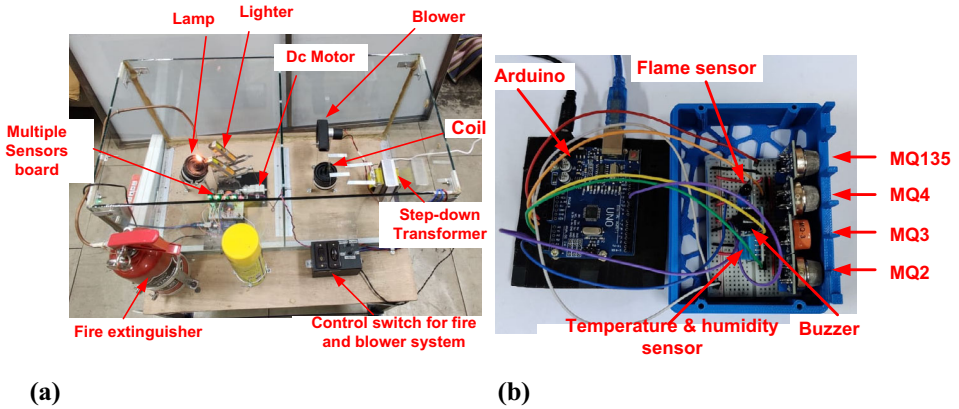


Figure 8. (a) Experimental laboratory test bed set up for fire detection system (b) Sensor node attached with different sensors.

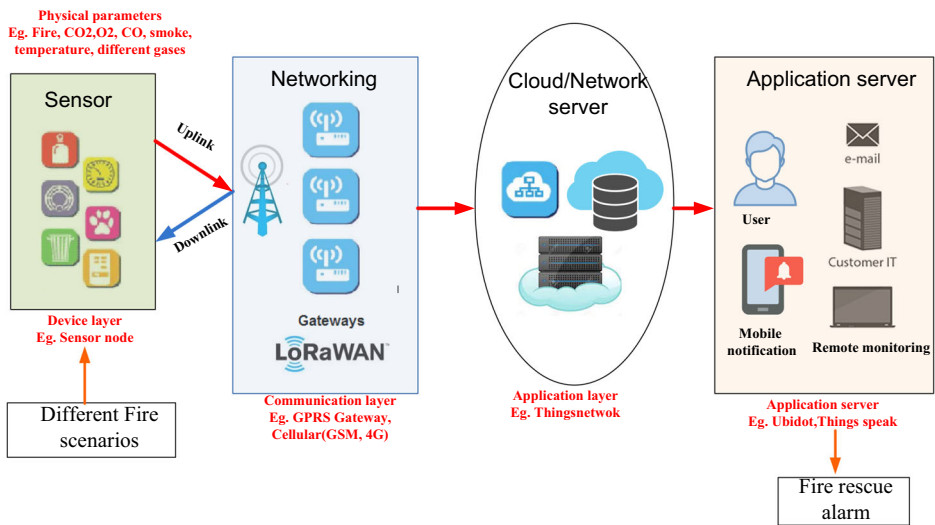
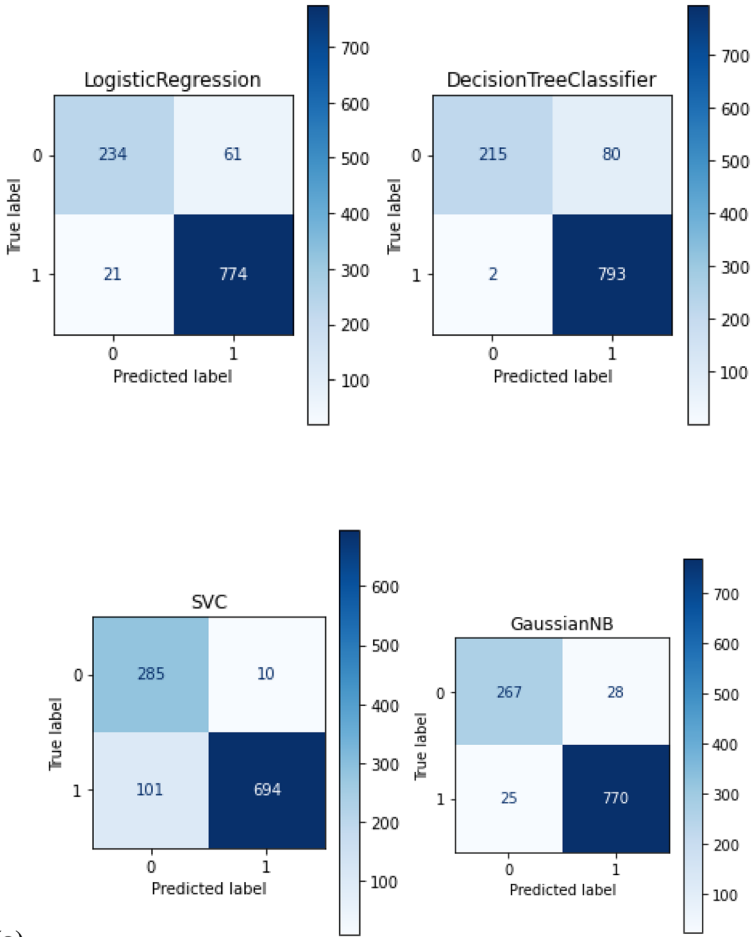


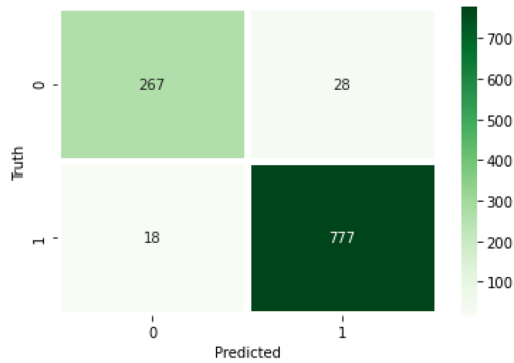
Figure 9. Illustrative IoT system floor plan and communication layer protocols for fire management.

Table 2 Confusion Matrices for Prediction Analysis

Actual	Prediction	
	Incorrectly predicted	Correctly predicted
Incorrect	T_P	F_N
Correct	F_P	T_N



(a)



(b)

◀ **Figure 10. (a) Confusion matrix plot for individual classifier (Logistic, Decision tree, SVM and Gaussian NB) (b) Prediction matrix for hybrid Ensemble classifier for fire detection system.**

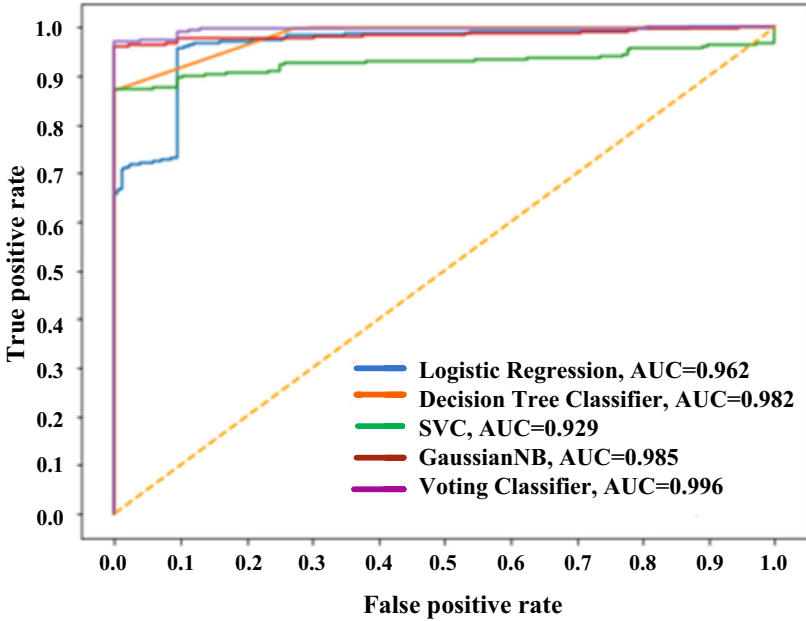


Figure 11. Roc plot of different classifier model and voting based ensemble model with AUC score.

4. Results And Discussion

4.1. Comparison of Confusion Matrix for Prediction

The predicted performance has been displayed by the confusion matrix plot which is a matrix array indicating the prediction condition compared with actual class. As we can see, the number of correctly predicted positive values and negative values are represented by T_P and T_N respectively. Accordingly, the number of incorrect classifiers is defined as F_P and F_N , as shown in Table 2. It is clear from the confusion matrix plot that the hybrid ensemble model performs very efficiently and classified the fire test data with a minimal error rate, as also has been seen in the confusion matrix in Figure 10.

The hybrid ensemble model is suitable for prediction of the majority classes of any problem and fails to predict the minority classes which are very challenging to perform real time application. Like most machine learning models, the proposed model also has a rate of misclassification, however, the rate of which is low. The reasons for the same are (a) High Bias—as a consequence when the model is ‘un-

der fitting' the training dataset of the example, and consequently, not presenting a very accurate relationship amongst the input and predicted variables.

(b) High Variance—due to a perfect fit of the proposed hybrid ensemble algorithm with the trained dataset. However, the developed model fits so well with the existing dataset that it may not give comparable results with the new sensor data, thus sacrificing accuracy. Instances of high bias can be solved by increasing the features in the data sets while high variances can be deal with by reducing sensitivity of the model by reducing features. An optimal balance of features has been considered after careful evaluation of the correlation matrix considering a sizeable dataset.

Based on the outcome of confusion matrix, accuracy can be defined as.

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N},$$

Where,

T_P = True positive slope correctly classified, T_N = True negative slope correctly classified, F_P = False negative slop incorrectly classified, F_N = False positive slope incorrectly classified.

Accordingly precision and Recall score is calculated defined as.

$$\begin{aligned} \text{Precision} &= \frac{T_P}{T_P + F_P} =, \text{Recall} = \text{TPR} = \text{Sensitivity} = \frac{T_P}{T_P + F_N}, \text{FPR} \\ &= \frac{F_P}{T_N + P_P} \end{aligned}$$

Where, TPR = It is defined correctly prediction positive, FPR = incorrectly predicted to positive.

The confusion matrix plot for hybrid ensemble method is shown in Figure 10 where where $T_P = 777$, $F_N = 267$, $T_N = 18$ and $F_P = 28$ and accuracy of the model have been tabulated in Table 2.

4.2. Comparison of the ROC curves and AUC Score

ROC analysis is a visual and numerical method used for distinguishing the given classes of classification algorithm and utilised for predicting structure and function from sequence data. ROC plot of individual learning model and hybrid ensemble classifier with average voting method is shown in Figure 11. A better classifier performance is observed when a particular ROC curve runs above the other ROC curve. With an AUC value closer to 1, better overall performance is noticed for final fire outbreak prediction by the proposed algorithm. AUC value of voting ensemble classifier is very closer to one suggestive of better prediction performance compared to other individual classifiers.

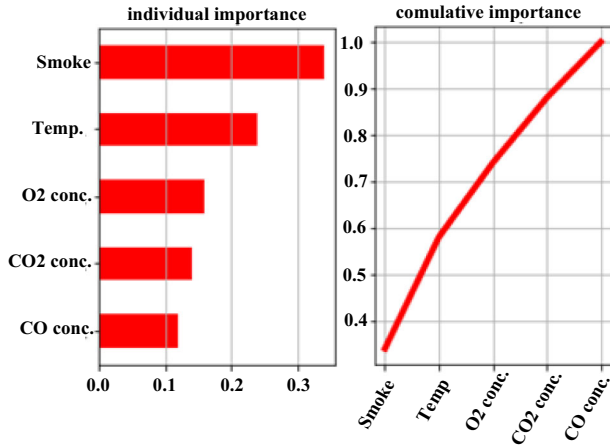


Figure 12. Feature importance—individual and cumulative plot for hybrid ensemble model.

4.3. Cumulative and Individual Importance of the Parameter of Hybrid Ensemble

Feature importance has been assigned a score for the respective features based on how useful it is in predicting a target variable and selection of that feature improves the efficiency and effectiveness of prediction of the problem. Individual importance and cumulative importance of the sensor dataset for hybrid ensemble classifier model is shown in Figure 12. The cumulative rising curve also helps us understand the relative weight of each of the contributing factors that are responsible for the fire detection.

Four individual classifier models have been trained and model scores have been recorded in performance evaluation Table 3, both for individual models and hybrid ensemble model. The hybrid ensemble model which is defined that each of the four individual machine learning models generates 4 times that results in a combination of a total of 20 weak learners of the model. After that, hybrid ensemble by average voting classifier technique is used wherein most of the classes have been predicted by the weak learner of the model may be the final prediction of the hybrid ensemble model. The model accuracy, AUC, precision, classification for prediction of the proposed hybrid ensemble model is seen to be better than the individual models with lower MAE, and RMSE error than individual classifiers. The performance of the proposed work with other similar research work for fire detection system has been illustrated in Table 4 where comparison is made in term of precision. It is observed that the performance of the proposed model is better than the existing approaches.

Table 3
Performance Comparison Table for Hybrid Ensemble Classifier Model with Other Single Classifier Model

Classifiers	Model precision	AUC score	Model score after train the model	MAE	RMSE
Proposed hybrid ensemble classifier with average voting technique	0.9752	0.996	0.98899	0.01009	0.10045
Logistic regression model	0.93809	0.962	0.93761	0.05688	0.23849
Decision tree model	0.93247	0.982	0.93027	0.0697	0.26405
Support vector machine	0.90947	0.929	0.90550	0.08532	0.29209
Naive bayes model	0.94983	0.985	0.94862	0.05137	0.22666

Table 4
Performance of the proposed work with other existing similar work

Approach	Precision	Method
Jiao et al. [67]	83	Deep learning (YOLOv3)
Lin et al. [68]	93.49	Contextual based
Jang et al. [69]	93.08	RF and threshold based
Fei Shi et al. [70]	0.778	Deep learning (YOLOv3)
Byoungjun et al. [71]	0.955	Deep learning
Proposed	0.9752	Deep learning (hybrid ensemble classifier with average voting technique)

5. Conclusion

In this manuscript, a hybrid ensemble model based on average voting technique is proposed for fire detection on real time multi sensor data. Four individual classifiers namely logistic regression classifier, support vector machine (SVM), Decision tree classifier and Naive Bayes classifier have been used which are seen to perform satisfactorily for fire detection. The proposed machine learning algorithm has been validated on five different fire scenarios and NIST dataset has been chosen for this purpose. The proposed ensemble classifier is observed to perform better than the constituent classifiers as well as reported literatures and results indicates improved model accuracy, AUC, precision with reduced Mean Absolute error, Mean Squared Error and RMSE error. Smart multi sensor fire detection device has also been developed which efficiently detects the presence of fire and wirelessly transmit the sensor data to cloud platform for further data analytics.

References

1. Bu F, Gharajeh MS (2019) Intelligent and vision-based fire detection systems: a survey. *Image Vis Comput* 91:103803
2. Wu H, Wu D, Zhao J (2019) An intelligent fire detection approach through cameras based on computer vision methods. *Process Saf Environ Prot* 127:245–256
3. Meacham BJ (1994) The use of artificial intelligence techniques for signal discrimination in fire detection systems. *J Fire Prot Eng* 6:125–136
4. Ko B, Cheong K, Nam J (2009) Fire detection based on vision sensor and support vector machines. *Fire Safety J* 44:322–329
5. Olivas JA (2003) Forest fire prediction and management using soft computing *Proceedings of the International Conference on Industrial Informatics (INDIN)*, pp. 338–344
6. Mahdipour E, Dadkhah C (2010) Automatic fire detection based on soft computing techniques: review from 2000 to *Artif. Intell Rev* 42(4):895–934
7. Anezakis VD, Demertzis K, Iliadis L, Spartalis S (2016) A hybrid soft computing approach producing robust forest fire risk indices. In: *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer International Publishing, pp. 191–203
8. Aertsen W, Kint V, Van J, Orshoven K., Ozkan, Muys B (2009) Performance of modelling techniques for the prediction of forest site index: a case study for pine and cedar in the Taurus mountains. *Turkey XIII World Forestry Congress*, pp. 18–23
9. Angelis AD, Ricotta C, Conedera M, Pezzatti GB (2015) Modelling the meteorological forest fire niche in heterogeneous pyrologic conditions. *PLoS ONE* 10(2):0116875
10. Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JM (2012) Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest *Ecol. Manage* 275:117–212
11. West AM, Kumar S, Jarnevich CS (2016) Regional modeling of large wildfires under current and potential future climates in Colorado and Wyoming USA. *Clim Change* 134(4):565–577
12. Bedia J, Herrera S, Camia A, Moreno JM, Gutiérrez JM (2014) Forest fire danger projections in the Mediterranean using ENSEMBLES regional climate change scenarios. *Clim Change* 122(1–2):185–199
13. Amatulli G, Camia A, San-Miguel-Ayanz J (2013) Estimating future burned areas under changing climate in the EU-Mediterranean countries *Sci. Total Environ* 450:209–222
14. Satir O, Berberoglu S, Donmez C (2015) Mapping regional forest fire probability using artificial neural network model in a mediterranean forest ecosystem. *Geomatics Nat Hazards Risk* . <https://doi.org/10.1080/19475705.2015.1084541>
15. Özbayoğlu AM, Bozer R (2012) Estimation of the burned area in forest fires using computational intelligence techniques. *Proc Comput Sci* 12:282–287
16. Yuan C, Zhang Y, Liu Z (2015) A survey on technologies for automatic forest fire monitoring detection and fighting using unmanned aerial vehicles and remote sensing techniques. *Can J Forest Res* 45(7):783–792
17. Denham M, Cortés AT , Margalef E (2008) Applying a dynamic data driven genetic algorithm to improve forest fire spread prediction *International Conference on Computational Science*. Springer Berlin Heidelberg, pp. 36–45
18. Bui DT, Bui QT, Nguyen QP, Pradhan B, Nampak H, Trinh PT (2017) A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agric For Meteorol* 233:32–44

19. Artés T, Cencerrado A, Cortés A, Margalef T (2016) Time aware genetic algorithm for forest fire propagation prediction: exploiting multi-core platforms. *Concurrency Computat: Pract Exper* . <https://doi.org/10.1002/cpe.3837>
20. Hong H, Naghibi SA, Dashtpajardi MM, Pourghasemi HR, Chen W (2017) A comparative assessment between linear and quadratic discriminant analyses (LDA-QDA) with frequency ratio and weights-of-evidence models for forest fire susceptibility mapping in China. *Arab J Geosci* 10:167
21. Bisquert M, Caselles E, Sánchez JM, Caselles V (2012) Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data. *Int J Wildland Fire* 21:1025–1029
22. Goldarag YJ, Mohammadzadeh A, Ardakani AS (2016) Fire risk assessment using neural network and logistic regression. *J Indian Soc Rem Sens* . <https://doi.org/10.1007/s12524-016-0557-6>
23. Maeda EE, Formaggio AR, Shimabukuro YE, Arcoverde GFB, Hansen MC (2009) Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. *Int J Appl Earth Obs Geoinf* 11:265–272
24. Safi Y, Bouroumi A (2013) Prediction of forest fires using artificial neural networks. *Appl. Math Sci* 7:271–286
25. Basheer I, Hajmeer AM (2000) Artificial neural networks: fundamentals computing design and application. *J Microbiol Methods* 43:3–31
26. Sakr GE, Elhadj IH, Mitri G (2011) Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Eng. Appl Artif Intell* 24:888–894
27. Xie DW, Shi SL (2014) Prediction for burned area of forest fires based on SVM model. *Appl Mech Mater* 513:4084–4089
28. Ko BC, Cheong KH, Nam JY (2009) Fire detection based on vision sensor and support vector machines. *Fire Saf J* 44:322–329
29. Zhao J, Zhang Z, Han S, Qu C, Yuan Z, Zhang D (2011) SVM based forest fire detection using static and dynamic features. *Computer Sci Inform Syst* 8:821–841
30. Wen T, Zhang BU. *University LT* (2014) Prediction model for open-pit coal mine slope stability based on random forest. *Sci Technol Rev* 32:105–109
31. Haifley T (2002) Linear logistic regression: an introduction. *IEEE International Integrated Reliability Workshop Final Report*, pp. 184–187 <https://doi.org/10.1109/IRWS.2002.1194264>
32. Rao P, Manikandan J (2016) Design and evaluation of logistic regression model for pattern recognition systems. *IEEE Annual India Conference (INDICON)*, pp. 1–6. <http://doi.org/10.1109/INDICON.2016.7839010>
33. Swain PH, Hauska H (1977) The decision tree classifier: design and potential. *IEEE Trans Geosci Electron* 15:142–147. <https://doi.org/10.1109/TGE.1977.6498972>
34. Navada A, Ansari AN, Patil S, Sonkamble BA (2011) Overview of use of decision tree algorithms in machine learning. *IEEE Control and System Graduate Research Colloquium, Shah Alam* . <https://doi.org/10.1109/ICSGRC.2011.5991826>
35. Zhao H, Yin S, Ru Z (2012) Relevance vector machine applied to slope stability analysis. *Int J Numer Anal Meth Geomech* 36:643–652
36. Yang Y, Jianping L, Yang Y (2015) The research of the fast SVM classifier method. *12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, pp. 121–124. <https://doi.org/10.1109/ICCWAMTIP.2015.7493959>
37. Mousa A, Ghasemian B, Shirzadi A, Shahabi H et al (2019) A novel hybrid approach of Bayesian Logistic Regression and its ensembles for landslide susceptibility assessment. *Geocarto Int* 34:1427–1457. <https://doi.org/10.1080/10106049.2018.1499820>

38. Chok YH, Jaksa MB, Kaggwa WS, Griffiths DV, Fenton GA (2016) Neural network prediction of the reliability of heterogeneous cohesive slopes. *Int J Numer Anal Meth Geomech* 40:1556–1569
39. Ren J, Lee SD, Chen X, Kao B, Cheng R, Cheung D (2009) Naive Bayes Classification of Uncertain Data. Ninth IEEE International Conference on Data Mining, Miami, FL, pp. 944–949, <https://doi.org/10.1109/ICDM.2009.90>
40. Dietterich TG (1997) Machine-learning research. *AI Mag* 18:97
41. Cho SB, Ryu J (2002) Classifying gene expression data of cancer using classifier ensemble with mutually exclusive features. *Proc IEEE* 90:1744–1753
42. Qi C, Tang X (2018) A hybrid ensemble method for improved prediction of slope stability. *Int J Numer Anal Methods Geomech* 42:1823–1839. <https://doi.org/10.1002/nag.2834>
43. Dogan A, Birant D (2019) A Weighted Majority Voting Ensemble Approach for Classification. 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 1–6, <https://doi.org/10.1109/UBMK.2019.8907028>
44. Jiakun Z, Ju J, Si C, Ruifeng Z, Bilin Y, Qingfang L (2020) A weighted hybrid ensemble method for classifying imbalanced data” *Knowledge-Based Systems*. ISSN 203:106087
45. He H, Garcia EA (2009) Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 21:1263–1284
46. Oscar C, Przemysław K, Bogdan T (2011) Special issue on hybrid and ensemble methods in machine learning. *New Gener Comput*. 29:241–244. <https://doi.org/10.1007/s00354-011-0300-3>
47. Wang N, Zhao S, Cui S, Fan W (2021) A hybrid ensemble learning method for the identification of gang-related arson cases. *Knowl-Based Syst* 218:0950–7051. <https://doi.org/10.1016/j.knosys.106875>
48. Gandhi Pandey M (2015) Hybrid Ensemble of classifiers using voting. *International Conference on Green Computing and Internet of Things (ICGCIoT)*, pp. 399–404, <http://doi.org/10.1109/ICGCIoT.2015.7380496>
49. Liu H, Gegov A, Cocca M (2015) Hybrid ensemble learning approach for generation of classification rules. *International Conference on Machine Learning and Cybernetics (ICMLC)* . pp. 377–382
50. Hsu K (2012) Hybrid ensembles of decision trees and artificial neural networks. *IEEE International Conference on Computational Intelligence and Cybernetics (CyberneticsCom)* 25–29, <https://doi.org/10.1109/CyberneticsCom.2012.6381610>
51. Yang S, Chen L, Yan T, Zhao Y, Fan Y (2017) An ensemble classification algorithm for convolutional neural network based on AdaBoost. *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. pp. 401–406
52. Lu H, Gao H, Ye M, Wang X (2019) A Hybrid ensemble algorithm combining ada-boost and genetic algorithm for cancer classification with gene expression data. *IEEE/ACM Trans Comput Biol Bioinf* 01:1–1
53. Saqlain M, Jargalsaikhan B, Lee JY (2019) A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. *IEEE Trans Semicond Manuf* 32:171–182. <https://doi.org/10.1109/TSM.2019.2904306>
54. Vahini Ezhilraman S, Srinivasan S et al (2019) Breast cancer detection using gradient boost ensemble decision tree classifier. *Int J Eng Adv Technol* 9:2249–8958
55. Renjie X, Haifeng L, Kangjie L, Lin C, Liu Y (2021) A forest fire detection system based on ensemble learning. *Forests* 12(2):217. <https://doi.org/10.3390/f12020217>
56. Sujatha G, Usha Rani K (2020) A Comprehensive Hybrid Ensemble Method with Feature Selection Techniques. *Advances in Computational and Bio-Engineering*. CBE

2019. Learning and Analytics in Intelligent Systems, vol.15. Springer, Cham. https://doi.org/10.1007/978-3-030-46939-9_8
57. Zhao J, Jin J, Chen S, Zhang R, Yu B, Liu Q (2020) A weighted hybrid ensemble method for classifying imbalanced data, Knowledge-Based Systems, 203. ISSN 106087:0950–7051. <https://doi.org/10.1016/j.knosys.2020.106087>
 58. Wang Y, Pan Z, Zheng J, Qian L, Li M (2019) A Hybrid ensemble method for pulsar candidate classification. Instrum Methods Astrophys. <https://doi.org/10.1007/s10509-019-3602-4>
 59. Verma B, Hassan SZ (2011) Hybrid ensemble approach for classification. Appl Intell 34:258–278. <https://doi.org/10.1007/s10489-009-0194-7>
 60. Chaudhary A, Kolhe S, Kamal R (2016) A hybrid ensemble for classification in multi-class datasets: an application to oilseed disease dataset. Comput Electron Agric 124:65–72. <https://doi.org/10.1016/j.compag.2016.03.026>
 61. Kardani N, Zhou A, Nazem M, Shen SL (2021) Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. J Rock Mech Geotech Eng 13(1):188–201. <https://doi.org/10.1016/j.jrmge.2020.05.011>
 62. Rosadi D, Andriyani W (2021) Prediction of forest fire using ensemble method. J Phys: Conf Ser 1918:042043
 63. Xie Y, Peng M (2018) Forest fire forecasting using ensemble learning approaches. Neural Comput Appl 31:4541–4550
 64. Stracher GB et al (2019) Gases generated during the low-temperature oxidation and pyrolysis of coal and the effects on methane-air flammable limits. In: Stracher GB (ed) Coal and peat fires: a global perspective Elsevier, Amsterdam, pp 157–171
 65. Nikunj C et al (2004) Ensemble Data Mining Methods NASA Ames Research Centre, USA
 66. Kohavi R (2001) A study of cross-validation and bootstrap for accuracy estimation and model selection. In International Joint Conference on Artificial Intelligence, pp. 1137–1143
 67. Jiao Z, Zhang Y, Xin J et al (2019) A deep learning based forest fire detection approach using uav and yolov3. In 2019 1st International Conference on Industrial Artificial Intelligence (IAI), Shenyang, China, 2019, pp. 1–5
 68. Lin Z, Chen F, Li B et al (2019) A contextual and multitemporal active-fire detection algorithm based on FengYun-2G SVISSR data. IEEE Trans Geosci Remote Sens 57(11):8840–8852
 69. Jang E, Kang Y, Im J, Lee DW, Yoon J, Kim SK (2019) Detection and monitoring of forest fires using Himawari-8 geostationary satellite data in South Korea. Remote Sensing 11(3):271
 70. Shi F, Qian H, Chen W, Huang M, Wan Z (2020) A fire monitoring and alarm system based on YOLOv3 with OHEM. In: Proceedings of the 2020 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020, pp. 7322–7327
 71. Kim B, Lee J (2019) A video-based fire detection using deep learning models. Appl Sci 9:2862

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.