MACHINE LEARNING APPROACHES USING PHOG AND SMOTE OF IMAGE FEATURE EXTRACTION TECHNIQUE FOR PREDICTING CONTROL VALVE STEM POSITION

B Kalaiselvi[1], Dr. Karthik B [2] Dr.A.Kumaravel [3]

Research Scholar, Dept of ECE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India

Associate Professor, Dept of ECE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India

Dean, School of Computing, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India

kalaigopal1973@gmail.com [1], karthikguru33@gmail.com [2], drkumaravel@gmail.com[3]

Abstract

Control valves are a progressively vital element of recent industrial instrumentation all over the world since these control valves are basically pneumatic devices composed of all mechanical parts their performance is less compared to the ideal one and due to constant moving parts wear and tear degrades over time. Control valves usually called the final control element in any process are used to control continuous flow, level, pressure, and temperature. The signal received from the conventional controller gives the signal for opening the CV (Control Valve) partially, fully open, or fully closed. The control signal is proportional to the magnitude of error with respect to time. The control valve is opened and closed automatically by giving the pressure to open and close using I/P converters in cascade with Pneumatic, Hydraulic, or electrical actuators with positioners. A plant can perform optimally if the performance of the control valve is monitored and maintained. The aim of this article is to initiate the studies in the direction of obtaining accuracy in stem positioning through image inputs and to demonstrate this yielding output as control valve stem position using machine learning algorithms, part of artificial intelligence enables identifying the stem position in less time using digital image processing. The proposed systems comprise of images of different stem positions is fed to the Weka software tool, preprocessed using Pyramid Histogram of Oriented Gradients (PHOG) feature filter and trained using pre-planned classifiers, the performance accuracy of stem levels by experiment-based attribute selection, iteration on ranking thresholds and SMOTE with triple iterations on class levels is calculated. The results show the performance with the maximum accuracy of 92.4051% and weighted average Receiver Operator Characteristics (ROC) values of 0.978. Hence such smart measurements using Machine learning algorithms which is a part of artificial intelligence provide us a vital role in predicting the CV stem position in less time using image processing filters.

Key words: Stem position prediction, Control Valve, FCE, Weka, PHOG Filter, Naïve byes, J48, Random Forest, and Instance-based Classification.

I Introduction

Control valve are also more often called a final control element in industrial automation. This plays a vital role in controlling the basic parameters like liquid level, pressure, flow for both Newtonian and non – Newtonian fluids. The conventional controller tuned in such a way that it compares the present output with the desired “set point” gives an output proportional to correct the deviation and obtain the steady state of the process.

A positioner with actuator in a control valve is used to linearize the stem position. The positioner will amplify the air supply from actuator to precisely open the valve for the given controller output. The signals being operated are followed universally like 4mA to 20mA, 3-15 psi or (1-5) Volts and So on but often (4-20) mA is used in industrial automation. In the current scenario a pneumatic control valve with positioner is used to collect the image data for building an optimum model using machine learning techniques. Traditional positioner may not integrate control typically but in this article the model is used to predict the stem position optimally and take necessary action thereby improving the closed control loop performances. The Control loop performance is affected by the movable parts in pneumatic valves due stiction and backlash effect are the two main problems that often occurs in FCE.

The module which gives me the position of the stem with the image fed into the module the output can also sent to get the controller output for driving the final control element. The 2 main failures in CV can be sorted out earlier by the proposed methodology. The smart measurements using Machine learning algorithms a vital part of artificial intelligence and gives us the stem position in less time of the process using image processing technique. Expected valve position is evaluated over a large vary of eventualities, multiple scenarios and therefore the absolute best acceptable performance at intervals the set threshold is looked for obtaining the ultimate best model.

Secondly the image data set non inheritable is preprocessed victimization real tool using SMOTE choice module. Training is taken place by activity classification over the information set chosen using a number of the classifiers associated the result's being tabulated for its performance accuracy and weighted average ROC. Finally, its performance analysis criteria for checking for higher performance and attainment of an optimum model. Moreover, this model has the aptitude of giving 3 completely different levels of the fluid by simply giving the image as input.

The convention method of running a 2 tank system is by default the process is a closed-loop system and the system is a second-order system. This is run using LabVIEW software by interfacing the module using a data acquisition system. The system is first executed in a run in auto mode and checking is done for any errors. Initially, the experimental check-up and all sensors are given the ready mode for running the process is achieved.

The performance evaluation criteria such as tr, td, Steady-state error are evaluated before the implementation of the optimum model. Through this research article, we would like to minimize the peak overshoots, fluctuations of the control valve early stage itself. This work also strengthens the idea of early prediction of stem position of this second-order system reduced.

This paper is organized into five sections as follows: The second section covers the related works associated with the proposed article supporting automatic CV stem position prediction and feature extraction and classification using machine learning algorithms. The third section contains a brief explanation of data preparation and description. The fourth section contains the proposed system flow and all the procedures for the validation of classification. The final section shows the output response of the attained model in the feedback of the proposed system block diagram.

II RELATED WORKS

This section shows that the related works of problem identification. Serrano, J et al. presented a computationally efficient machine learning model for fault detection in an mDSF engine using a three-class Logistic Regression solution based on commonly available engine controller signals. Training and testing accuracy exceeded 98% based on steady-state engine dyno data with valve faults induced at a 1% rate. [1]

Dangut et.al. [7] Proposed model is compared to other similar deep learning approaches. The results indicated an 18% increase in precision, a 5% increase in recall, and a 10% increase in G-mean values. It also demonstrates reliability in anticipating rare failures within a predetermined, meaningful time frame.[8-16].

Lawrence et.al. [17] the deployment of Linear Regression as a Machine Learning technique for prediction of cavitation based on observational data collected from sensing instruments monitoring the process condition and the control valve under study. The machine learning algorithm implemented for the identification of various failures in the control valve and can say that the predictive analysis is more efficient than fuzzy logic as the fuzzy logic requires data to be collected for all the failure events and define membership functions for each type of failure.[18-21]. By using a machine learning approach during the operation of a well with multiple ICV settings, it would be feasible to estimate the lateral-by-lateral output at unseen scenarios. Hence, it becomes possible to maximize the well output by using an optimization algorithm to determine the optimal ICV settings.[22]

III DATA PREPARATION AND DESCRIPTION

The below lying information happens to be a set of pictures showing essentially the Stem positions with variable lighting background. The image dataset obtained therefore inevitably contains imprecise, strident pixels to be regulated to preserve the mensuration accuracies. Stem position measurement is finished by employing a completely different approach to victimization image processing.

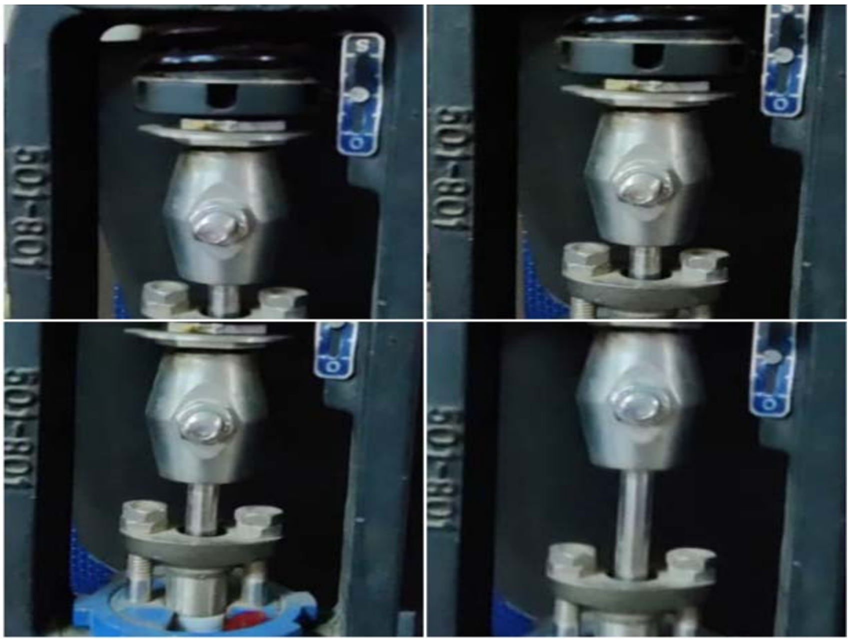

Figure 1 Final control Element Scenario for Stem position

Figure 1 shows the physical appearance of the 2-tank system with the control valve in the process station the data collection is done with this two-tank system and the different stem level images are acquired through a normal high-resolution camera. The collected images are labeled with its actual percentage of stem position and then modeling the 2-tank interacting tank system which is a second-order process system is done using the traditional method of running the process with mandatory controllers like PI, PID controllers with tuning parameters, and a steady-state is obtained after some prescribed interval of time.

The images of the control valve for various positions are obtained and are tabulated, labeled for implementing machine learning algorithms using supervised learning. The data collection is done based on the stem positions as shown in figure 2 which describes the different stem positions as 0%, 5%, 50% and 100% openings of the control valve during the running of the process shown in figure 2.

Using digital image processing and image filters like PHOG are used for preprocessing the image to the corresponding numeric values for classification of different classes in machine learning algorithms like rules bases, nearest neighborhood, lazy rules and Trees based algorithms.

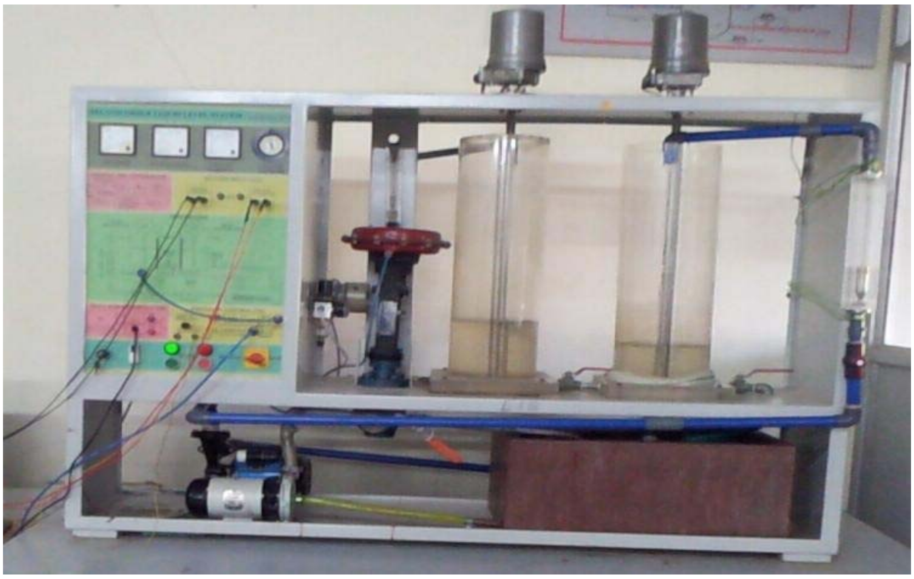

Figure 2 Two interacting tank system set up

3.1 Mathematical modeling of two-tank system

The convention method of running the process with the general parameters and observed for the self-regulatory control of the system. The open-loop response with the given setpoint of 12 cm and the default controller's parameters are set and the process is run till it reaches the stable state that is plus or minus 5% of the set point.

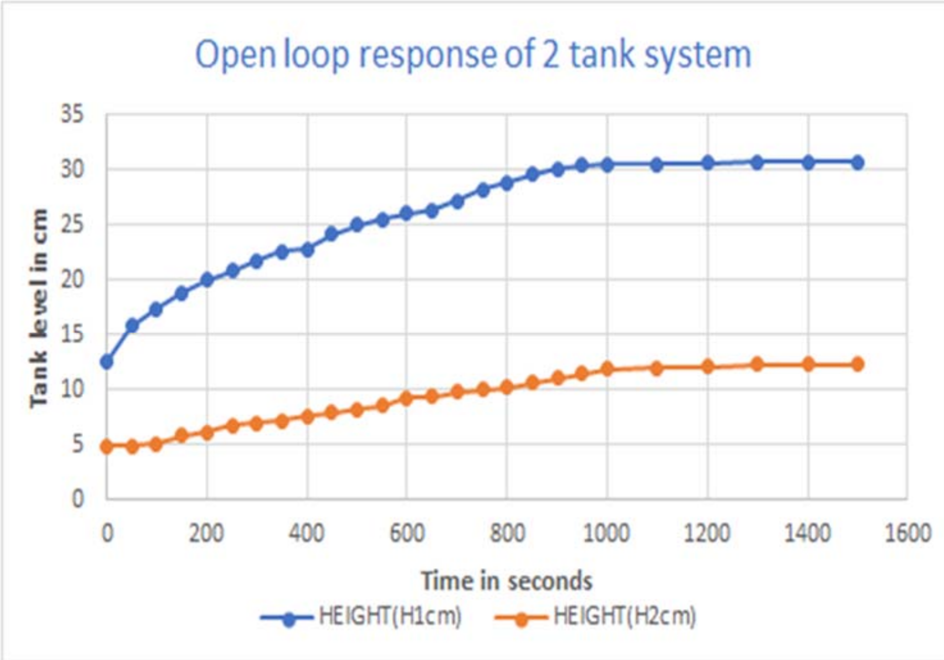

Figure 3 open response of 2 interacting system

From the above, figure 3 it is observed that system reaches the stable state for the given set point for thousandth second in open loop condition. Mathematical modeling is done by finding the values of the time constants τ1 and τ2 by calculating the product of the rate of accumulation and resistances R_1 and R_2 at 63% of its final value. However, the values of R_1 and R_2 are evaluated to be 254,77 and 339.49. The resultant transfer function by substituting the values is:

$$\frac{H2(S)}{Qi(S)} = \frac{339.49}{85.28S2+29.32S+1}$$ ---2nd order equation

The figure 5 illustrates the Simulink model of the above-mentioned open loop system. Simulated with the obtained transfer function and giving a step input and seeing the corresponding output in the scope.

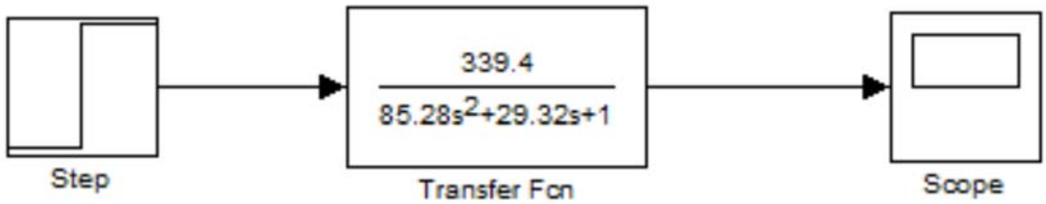

Figure 4 Simulink diagram for open loop

The output of this Simulink for completing the mathematical modeling is given in the below figure 4 for a step input to the obtained transfer function is:

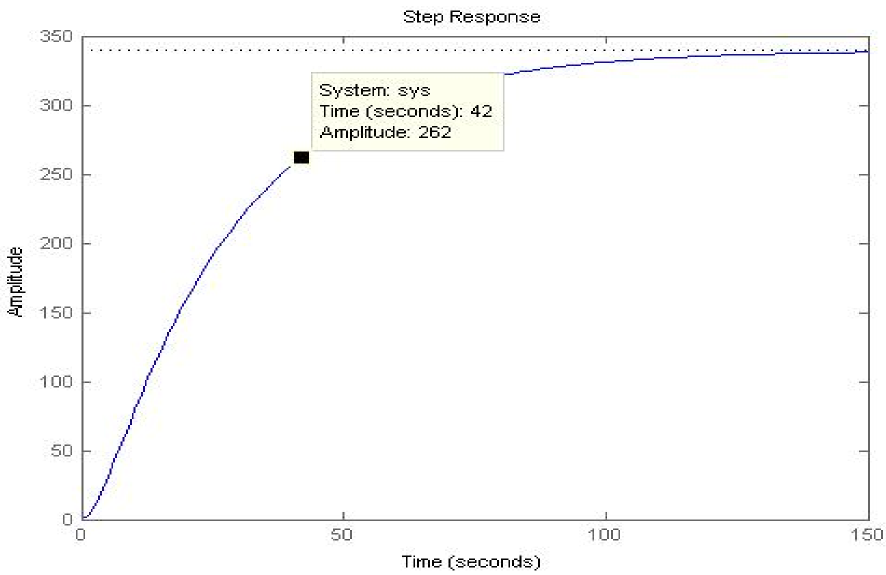

Figure 5 Step input response to the Transfer function obtained.

Figure 5 is the response achieved for the obtained transfer function, which is the time versus output amplitude characteristics. The process control system is run in the conventional method to get the image data of the control valve to progress to build a model in feedback to predict the control valve stem position using machine-learning algorithms.

The Control valve images are collected by running the process with the default controllers and the CV images are collected under three different classes and tabulated by labeling the level values. The true value of the collected CV images is compiled in the given format with the output CV stem positions. For example, CV1.jpeg = Low, CV6.jpeg = Medium, CV14.jpeg = High and so on for all the 58 distinct images obtained. Then the file format must be changed for further processing. The filter section is used for which the file format supported is ARFF format i.e., Attribute relation file format, combining the image and the class in one single file and converting it into a single file for further processing.

3.2 Major seven steps in data preprocessing are:

- Acquire the dataset
- Import all the crucial libraries
- Import the dataset
- Identifying and handling the missing values
- Encoding the categorical data
- Splitting the dataset

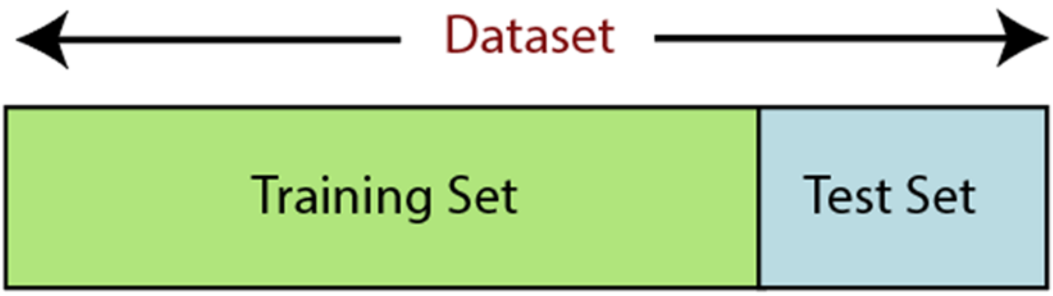

Figure 6 Partitioning the dataset during training

- Feature scaling

Most ML models are based on Euclidean Distance, which is represented as: $d(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

Figure 7 Feature extraction Pre-processing steps

Number of features: 631, The data type is Numerical with 58 instances and 3 classes with specified set points as maximum statistic value as 8 and minimum statistic value as 0 and Standard deviation 1.681.

Data preprocessing is a have to. There are 3 approaches to inject the records for preprocessing:

• Open File – permits the consumer to pick the document from the nearby gadget

• Open URL – permits the consumer to pick the records document from one-of-a-kind locations

• pen Database – permits customers to retrieve a records document from a database source

Data preprocessing is carried out in two different stages for converting the image data into numerics using PHOG image filter method and SMOTE for balance the imbalance data set for the applied classes that will affect the performance of the model built to predict the stem position according to the level flow and the desired setpoint of the control valve.

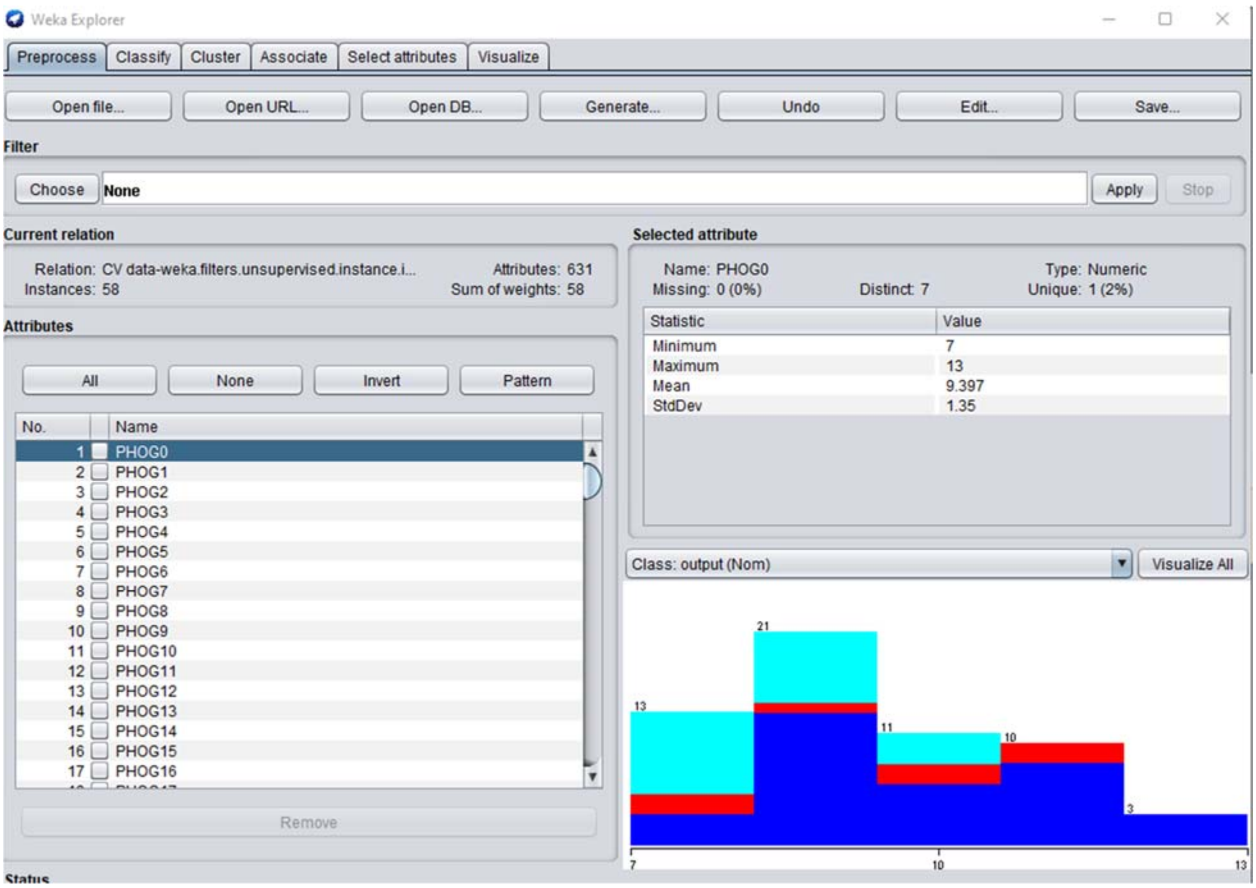

Figure 8 Weka based PHOG data feature extraction preprocessing data

SMOTE (Synthetic Minority Oversampling Technique)

SMOTE is used to overcome the imbalanced data set to improve the accuracy performance using Weka software. SMOTE is a common strategy for dealing with unbalanced classes in classification problems. SMOTE as shown in figure 8 is an oversampling technique where the synthetic samples are generated for the minority class. If the unbalanced data is not taken care of beforehand it may degrade the performance of the classifier model. Most of the predictions will be made with the majority class wherein the minority will be treated as noise and disturbance and will be ignored by the classifier algorithm resulting in a high bias in the model.

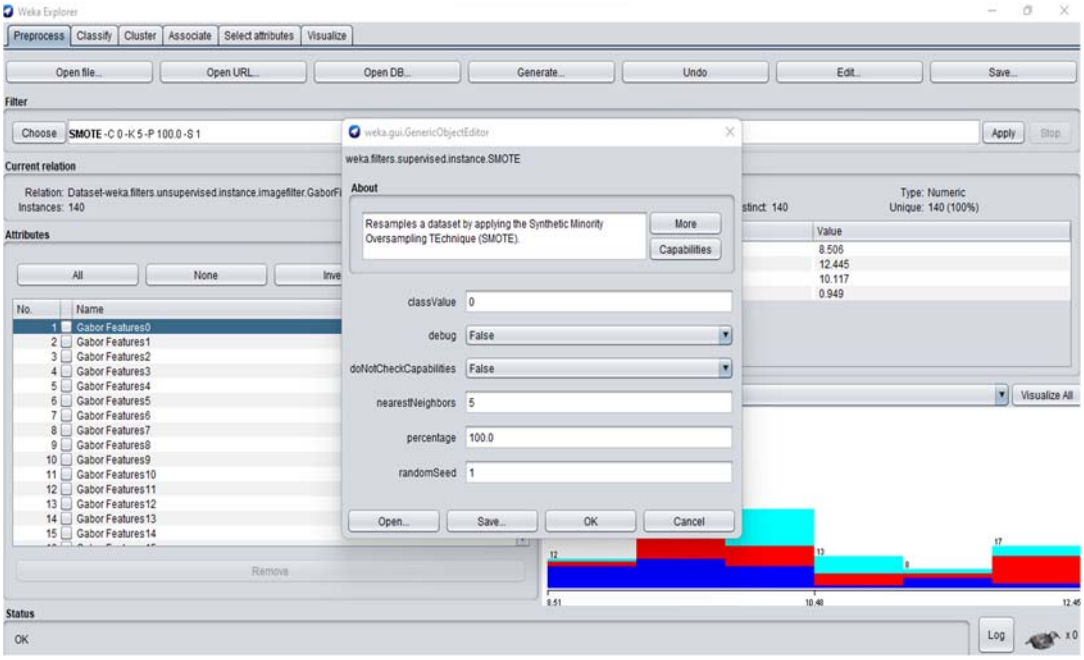

Figure 9 Weka based image data feature extraction using SMOTE filter

Option panel for SMOTE filter is shown in figure 9 that can be carried out in different percentage of sampling the minority case. The option adopted varied fro 50% to 100% of oversampling the minority class in the dataset.

In our article the SMOTE iteration is carried out in three different stages and tabulated as shown in the table1 for performing nearly two times 100% minority sampling. Two times the minority class is sampled that the minority class instances will be doubled as shown in the look up table1.

Table 1 Lookup image data set with and without applying SMOTE

CV Stem position Class	**Instances count for first iterated model**	**Instances Count @SMOTE1 for Second iterated model**	**Instances Count @SMOTE2 for third iterated model**
Low	33	33	33
Medium	7	14	28
High	18	18	18

Table 1 comprises of describing the three different classes as High, Medium, and Low having the instances count of 18,7, and 33. The High class specifies the stem position greater than the 50%, the medium class is around the 50% with 5% tolerances, and the Low class specifies the CV stem position below 50%. Applying SMOTE is resampling the minority class for three iterations is also specified in table 1.

3.3 Block Diagram of 2 Tank System with Smart Positioner

The block diagram clearly establishes the optimum model in the closed-loop for governing the system by measuring the stem position optimally as shown in figure 10. This also shows some enhancement in the earlier prediction of stem position and thereby maintaining the plant to run continuously without any breakdown and early prediction of failures like backlash and stiction problems.

It is a cascade loop having primary and secondary loop with two controllers one monitoring the flow and the other the level in the process. The secondary controller undergoes the problem of continuously varying setpoint and tuning a controller for settling the process becomed very difficult task hence implemented an optimum controller using machine learning algorithms

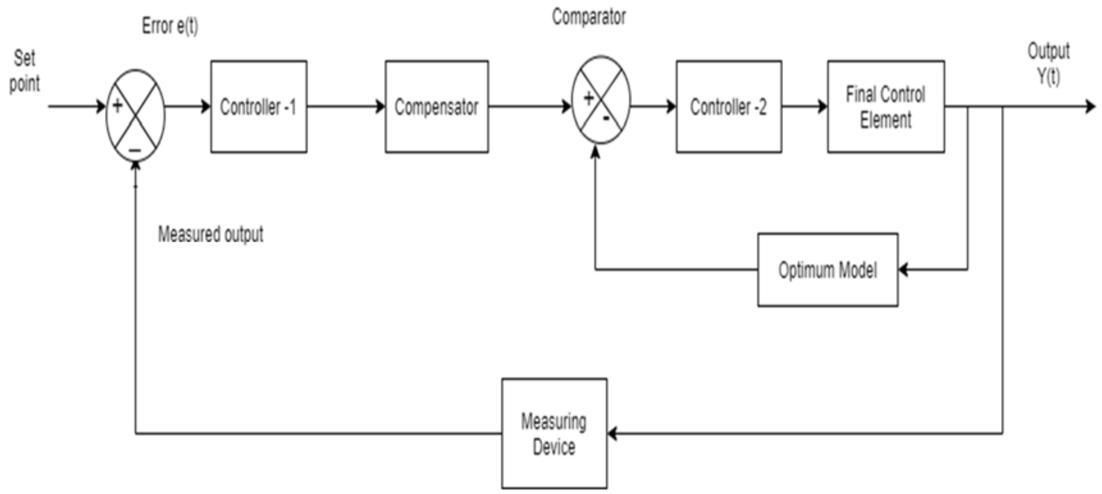

Figure 10 Designed optimum model in the feedback

IV Proposed System flow for Smart Positioner for Control Valve

WEKA or Waikato Environment for Knowledge Analysis the University of Waikato Hamilton, New Zealand is an open-source data mining software issued under GNU public license software that provides freedom to its users in performing all these data mining works.[22]

It is a collection of many machine learning algorithms for data mining tasks. It is a pack of a tool consisting of various operations called document preparation, clustering, regression, data preprocessing, classification, Association rules, instance-based classifying, and picturing. The methodology involved in establishing the task is carried out using this software called WEKA for modeling the above said smart positioning using image processing.

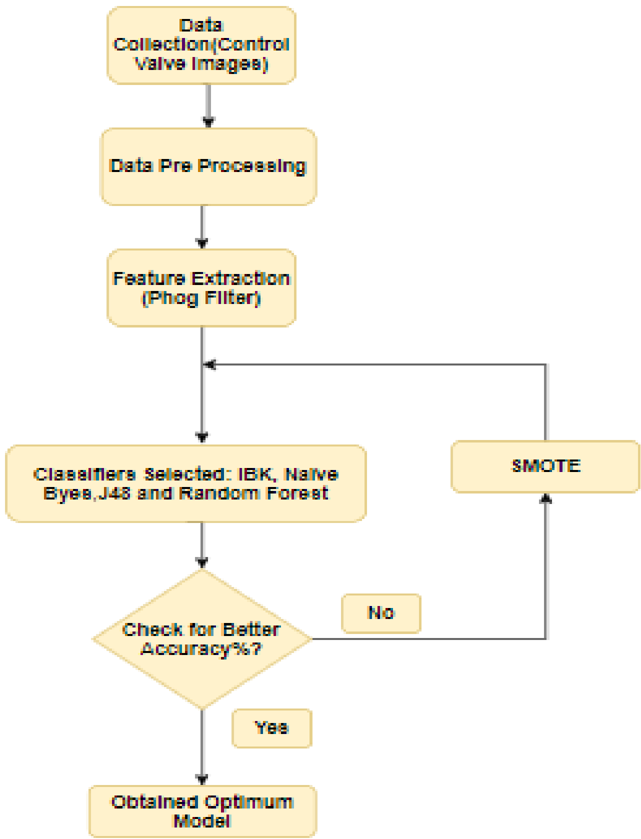

Figure11 Experimental set up to obtain Optimum model

Figure 11 shows the experimental set up for predicting the Control valve stem position in the following steps:

- Data collection as raw images that has to cropped, format changed and framed the ground truth values to the corresponding images
- Data preprocessing using PHOG filter and the image to numeric values is processed for further classification and building the model using selected classifiers
- Classified and built the model for training and obtained performance results
- Inferring the results for better accuracy percentage and area of curve called weighted receiver operator characteristics is better or not for repeating the above process by applying SMOTE to the minority class to improve the overall performance of the built model.
- Attainment of the optimum model to predict the control valve stem position using machine-learning algorithms.

The control valve images for various stem positions namely 0%,25%, 50%,75%, and 100% are formed into a lookup table and converted into a comma-delimited file in excel, and the data preprocessing, feature extraction is carried out using PHOG filter in weka software in functions tab. Then the classifying process is done using the classifiers namely instance base classifiers, Naive Byes, Decision tree classifiers(J48, Random Forest). Tabulating the results and checking for its accuracy level percentage. A synthetic minority imbalance in the dataset is overcome by a tool called SMOTE to get better accuracy performance. This SMOTE will give us better evaluation criteria in performance accuracy by balancing a low-resolution image also.

Description of preprocessing using SMOTE image filter: Synthetic Minority Oversampling TEchnique common strategies for dealing with unbalanced class in classification problems. SMOTE is an oversampling technique where the synthetic samples are generated for the minority class. If the unbalanced data is not taken care of beforehand it may degrade the performance of the classifier model. Most of the predictions will be made with the majority class wherein the minority will be treated as noise and disturbance and will be ignored by the classifier algorithm resulting in a high bias in the model. SMOTE is used to increase the classifier performance despite the imbalance in the data set. The effect of 'SMOTE' clearly shows a uniform increment in the decision rule classifier [21].

4.1 Procedure for validation of classification: The preprocessing is done with a PHOG filter with 631 attributes and 58 distinct instances, having a mean value of 9.397 and Standard deviation equals 1.35

PHOG Filtering: Pyramid Histogram of Oriented Gradients (PHOG) features have been employed to discriminate between different stem positions. The proposed shape discriminator counts occurrences of gradient orientation in the localized portion of an image.

To improve the recognition rate Pyramid Histogram of Oriented Gradients (PHOG) is used for feature extraction. As an example of the feature after filtering by PHOG scheme, its (PHOG10) distribution is shown in Figure12.

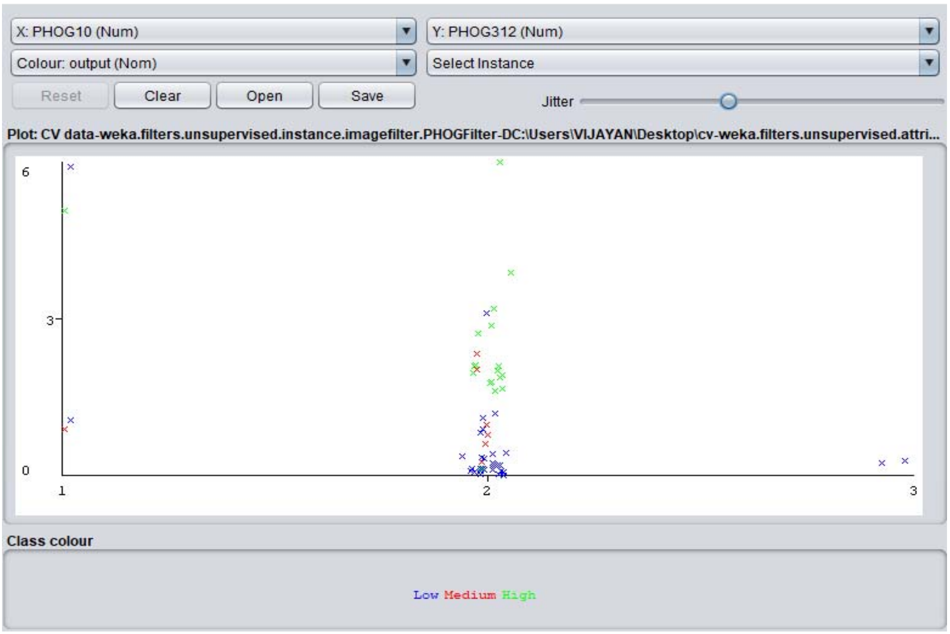

Figure 12 Pre-Processing using image Filter (PHOG)

The PHOG features are mined from the given input (Control valve stem Position) related to the stem position. It is basically a three-dimensional shape descriptor applied to image sorting. The PHOG features are mined from the region of interest (Control valve stem position) are focused by its local feature that is captured over edge positioning within a region and three-dimensional layout of the image.

The three-dimensional distribution of edges is formulated as a vector image by tiling the image into a region at multiple resolutions. Pyramid histogram of orientation gradients over each image sub-region at each resolution.

4.2 Procedure for selected base classifiers:

- Naïve Bayes
- IBK instance base Classifiers(KNN)
- Random Forest
- J48

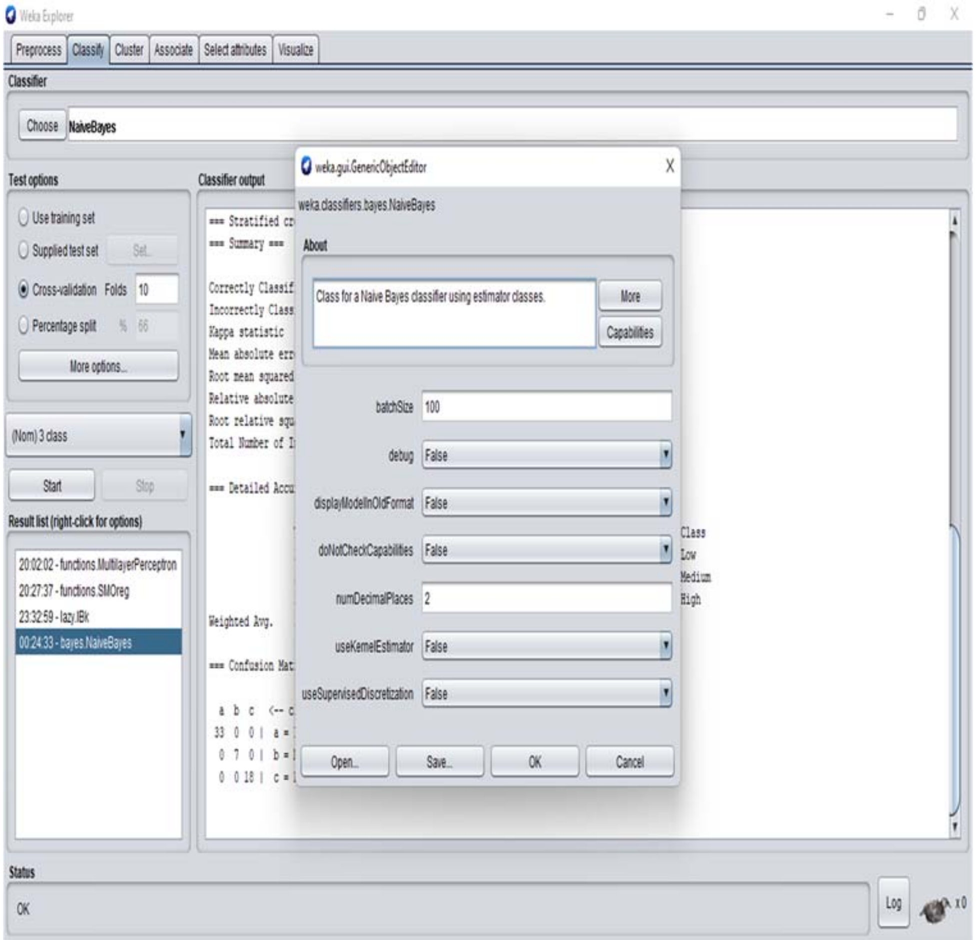

Figure 13 Weka parameter tuning option panel for Naïve Bayes Classifier

Bayes' theorem (BT) is defined as the probability of an event that occurs in the given probability of another event which has already occurred. BT is stated mathematically:

$$P\left(\frac{A}{B}\right) = \frac{P(B/A)P(A)}{P(B)}$$

Where A and B are the events and P(B) not equals to 0. In this regard the dataset is applied BT in the following way

$$P\left(\frac{y}{X}\right) = \frac{P(X/y)P(y)}{P(X)}$$

Where y is the class variable and X is a dependent feature vector with n elements: $X = (x_1 + x_{2+} x_{3+} x4_{\ldots\ldots+} x_n)$

Figure 13 shows the Weka parameter tuning option panel for Naïve Bayes Classifier which is in the Bayes category in Weka.

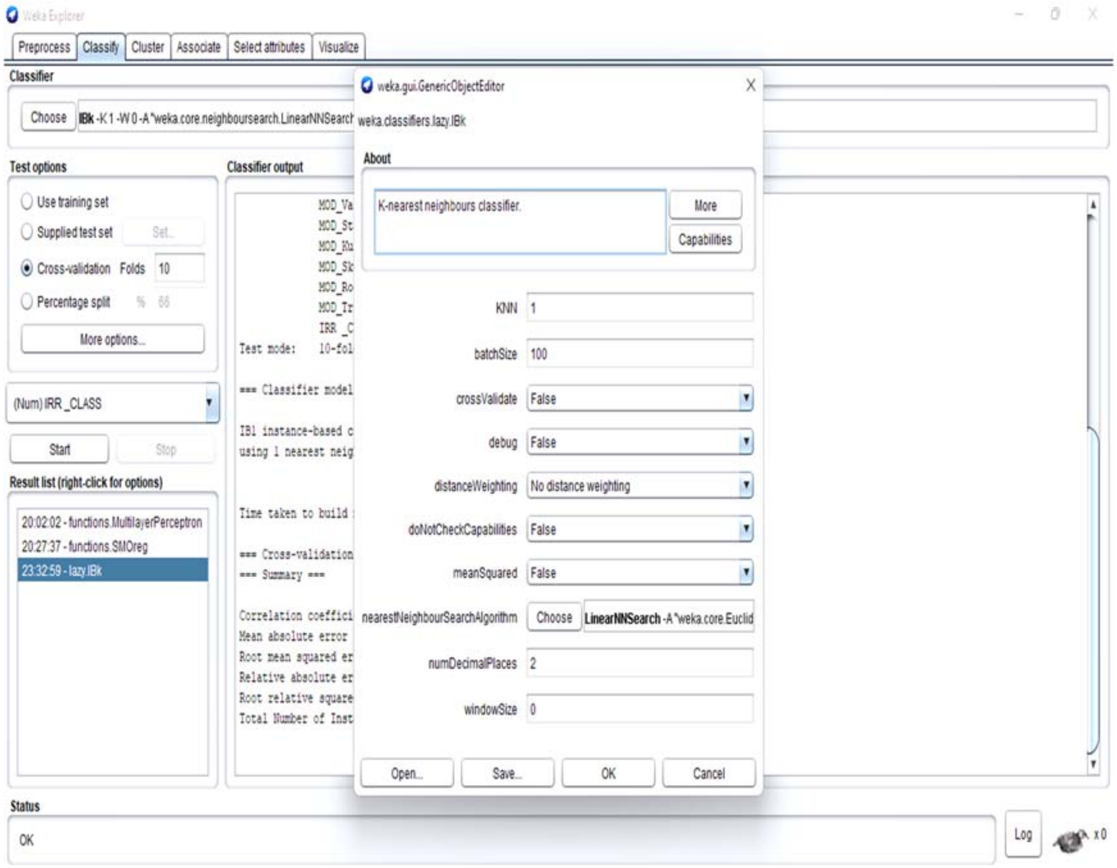

Figure 14 KNN Classifier in parameter tuning option panel

For any point of data y, the Euclidean Distance from another data point x is given by ,

$$ED = \sqrt{\sum_{i=1}^{n}(x_i + y_i)}\ \hat{}2$$

where n is the number of features.

Figure 14 KNN Classifier in parameter tuning option panel in Weka and this classifier has opted lazy category and the option panel is utilized for varying the neighbor hood values.

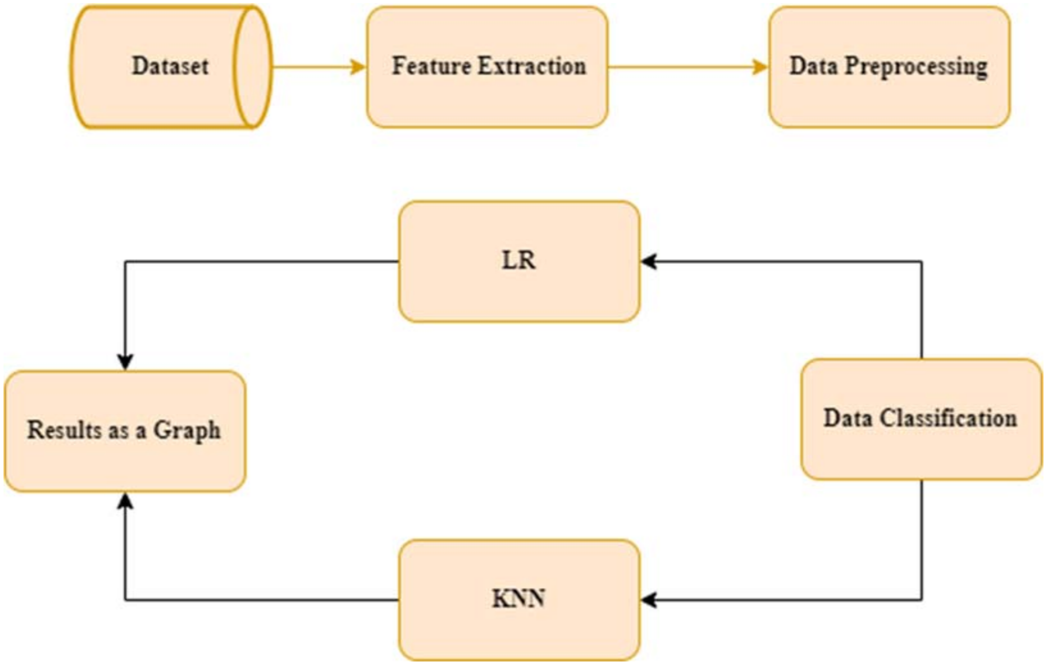

Figure 15 Analysis of KNN algorithm

The value of neighbors that is k value is chosen depending upon the performance as shown in figure 15, required accuracy, and the correlation Coefficient accepted for the built model. The advantage of this KNN algorithm is it is very simple to implement, Robust to the noisy training data, and it can be more effective on the training data of very large size.
The drawback of this algorithm is it always needs to determine the value of k which may be Complex in some cases, the computational cost is high because of calculating the distance between the data points for samples.
Random Forest classifier and J48 are from tree based category and the implementation procedure is as follows:

- Data preprocessing procedure
- Fit the Random Forest algorithm to the training set
- Prediction of test results
- Test the accuracy of the result (creating a confusion matrix)
- Visualization of test series results.

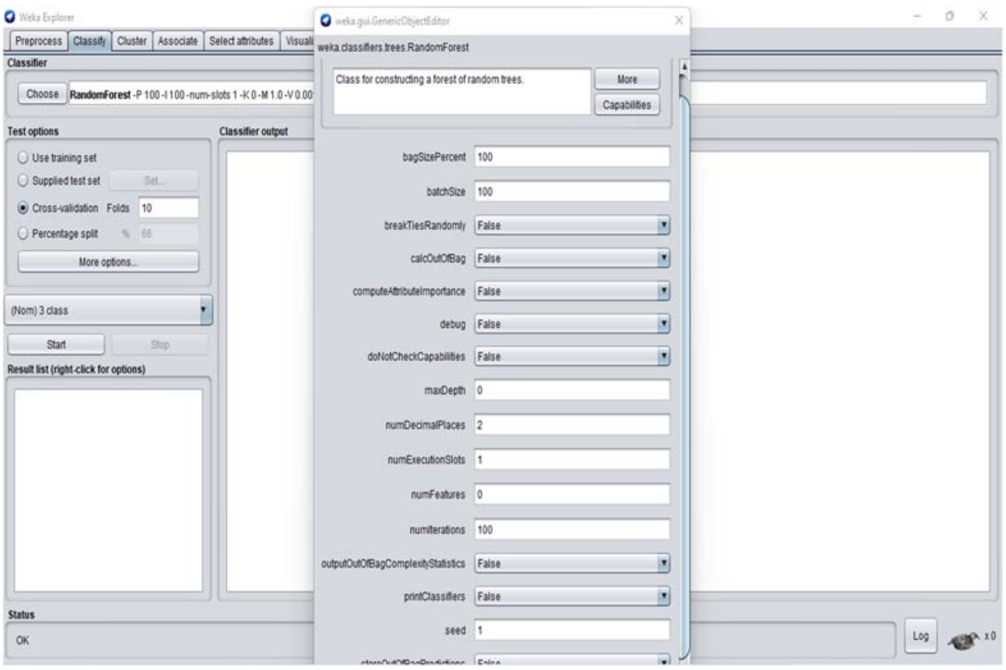

Figure 16 Option panel for parameter tuning Random Forest classifier

- The general capability of the Random Forest classifier is the Binary class, Missing class values, and Nominal class.
- The attribute that it uses is Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, and Unary attributes.

The classifier object takes the following parameters as shown in figure 16 :

n_estimators = Number of trees required for Random Forest. The default value is 10. You can choose any number, but you have to deal with the problem of overfitting.

Criteria = A function that analyzes the accuracy of division. Here, we used "entropy" to get the information.

The need for confusion matrices in machine learning

- Evaluate the performance of the classification model when making predictions about test data and show how good the classification model is.
- It shows the nature of the error as well as the error caused by the classifier. B. Is it a Type I or Type II error?
- You can use the confusion matrix to calculate various parameters of your model, such as accuracy and accuracy.

Calculations using Confusion Matrix:

We can use this matrix to perform various calculations on your model, such as model accuracy. These calculations are shown below.

Classification accuracy: This is one of the important parameters to determine the accuracy of the classification problem. Defines how often the model predicts the correct output. This can be calculated as the ratio of the number of correct predictions made by the classifier to the total number of predictions made by the classifier. The formula is:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + FP + FN + TN}}$$

- Where TP and TN are the True positive and True Negative.
- FP and FN and False positive, False Negative Classification

Misclassification rate: Also known as the error rate, it defines how often the model makes false predictions. The error rate value can be calculated as the number of false predictions out of all the predictions made by the classifier. The formula is:

$$\boldsymbol{Error\ rate} = \frac{\boldsymbol{FP + FN}}{\boldsymbol{TP + FP + FN + TN}}$$

Table 2 Experimental results of the first iterated model

S.NO	**Category**	**classifier**	**Accuracy (%)**	**Weighted Avg ROC**
1	Bayes	Naive Bayes	82.7586	0.850
2	Lazy	IBK	79.3103	0.816
3	Trees	J48	69.9655	0.763
4	Trees	RF	82.7586	0.952

Table 2 results clearly show the classification, of the stem position of the final control element is correctly classified by the above-mentioned classifiers are above 69%accuracy and the Receiver operator characteristics is also having a weighted average above 0.763 and the approximate time taken to build the models is also very less.

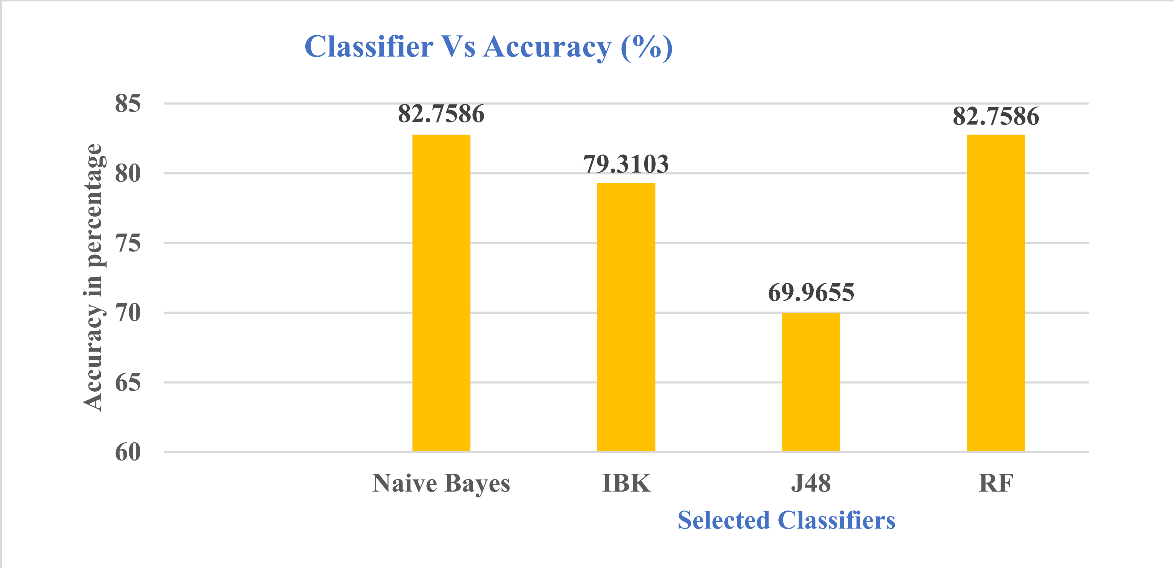

Figure17 Classified data performance analysis1

We observed from figure 17 the accuracy starts from 69% to 82.7586% for J48 and Naïve Bayes classifier, Random Forest.

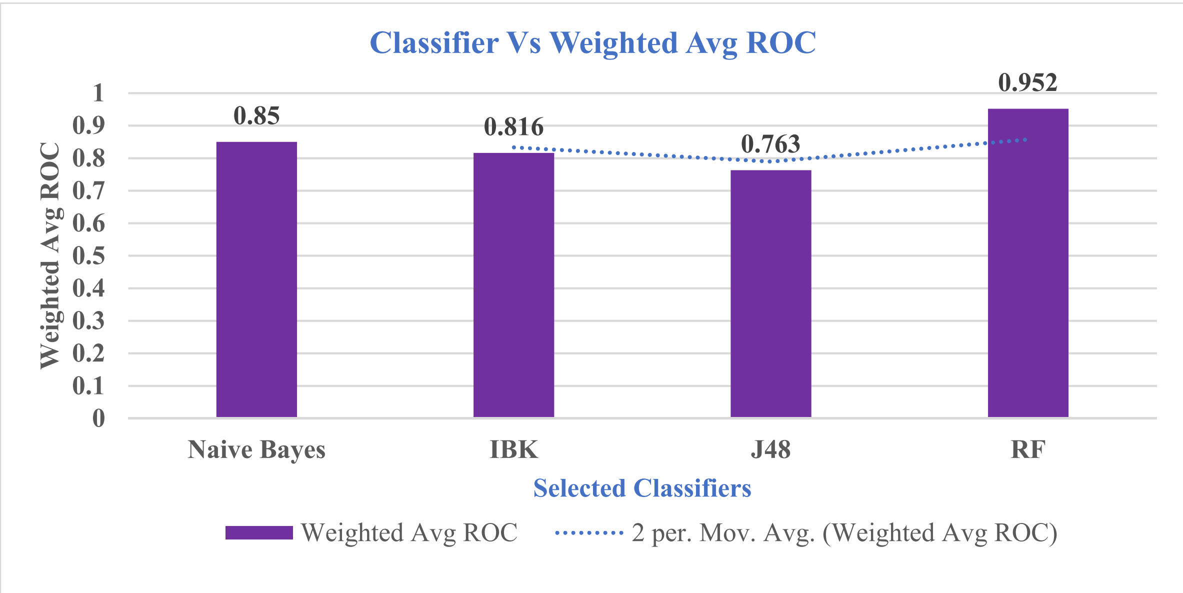

Figure 18 Classifier Vs Receiver Operator Characteristics1

ROC is observed from figure 18 varies from 0.796 to 0.952 for J48 and Random forest, both belonging to the same tree category.

The second iterated model is built by preprocessing using SMOTE (Synthetic Minority Oversampling TEchnique) by increasing the data of minority by in Weka tool followed by the classification is done.

Table 3 Experimental results of the Second iterated model

S.NO	Category	classifier	Accuracy (%)	Weighted Avg ROC
1	Bayes	Naive Bayes	80	0.821
2	Lazy	IBK	89.2308	0.919
3	Trees	J48	69.2308	0.796
4	Trees	RF	87.6923	0.952

The experimental results of the second iterated model from table 3 show a decrease in accuracy performance for Naïve Bayes and an increasing percentage for the remaining IBK, J48, and Random Forest algorithms. The selected classifier versus Accuracy performance is shown in figure19. The ROC for the selected classifiers also observed increasing area attained as shown in figure 20.

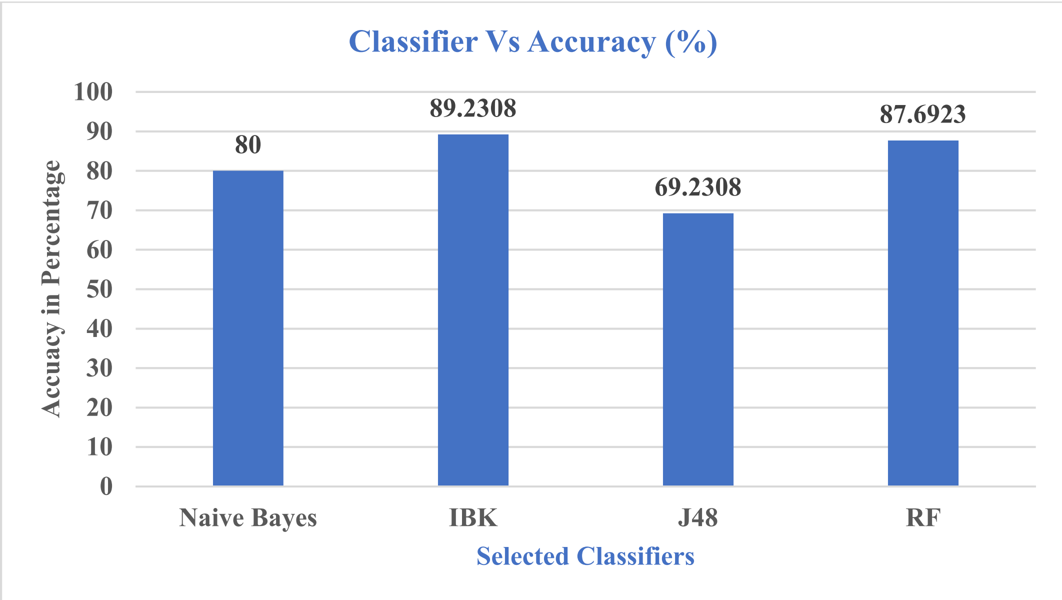

Figure 19 Classified data performance analysis2

The Second iterated model is built using preprocessing using SMOTE in Weka which is used for increasing the classifier performance despite imbalance data. We detect the accuracy starts from 69% to 89.2308% for J48 and Naïve Bayes.

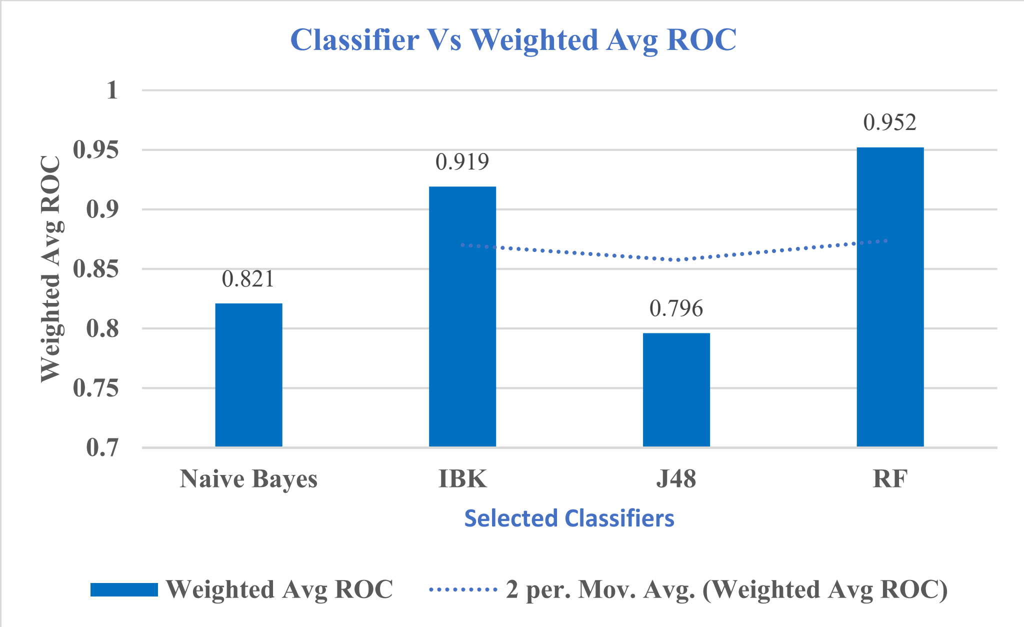

Figure 20 Classifier versus Receiver Operator Characteristics2

The ROC in figure 20 shows a weighted average of 0.952 for Random Forest.

The third iterated model is built using preprocessing using SMOTE in Weka which is used for increasing the classifier performance despite imbalance data as shown in table 4.

We detect the accuracy starts from figure 21 shows 78.481% to 92.4051% for J48 and Random Forest.

Table 4 Experimental results of the third iterated model

S.NO	Category	classifier	Accuracy (%)	Weighted Avg ROC
1	Bayes	Naive Bayes	87.3418	0.912
2	Lazy	IBK	86.0759	0.881
3	Trees	J48	78.481	0.827
4	Trees	RF	92.4051	0.978

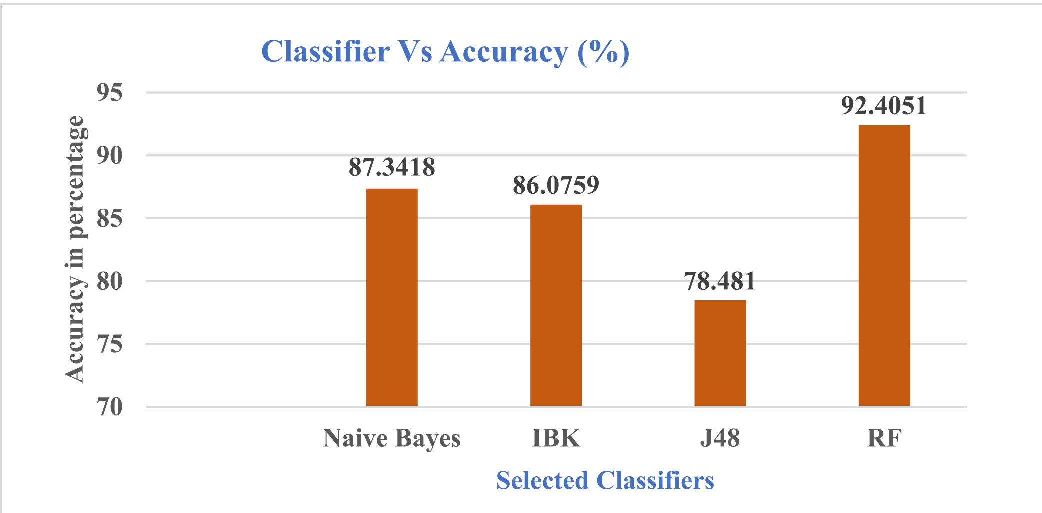

Figure 21 Classified data performance analysis3

The ROC gives a weighted average of 0.978 as shown in figure 22. With these results, we conclude that we have attained an optimum model in the feedback loop.

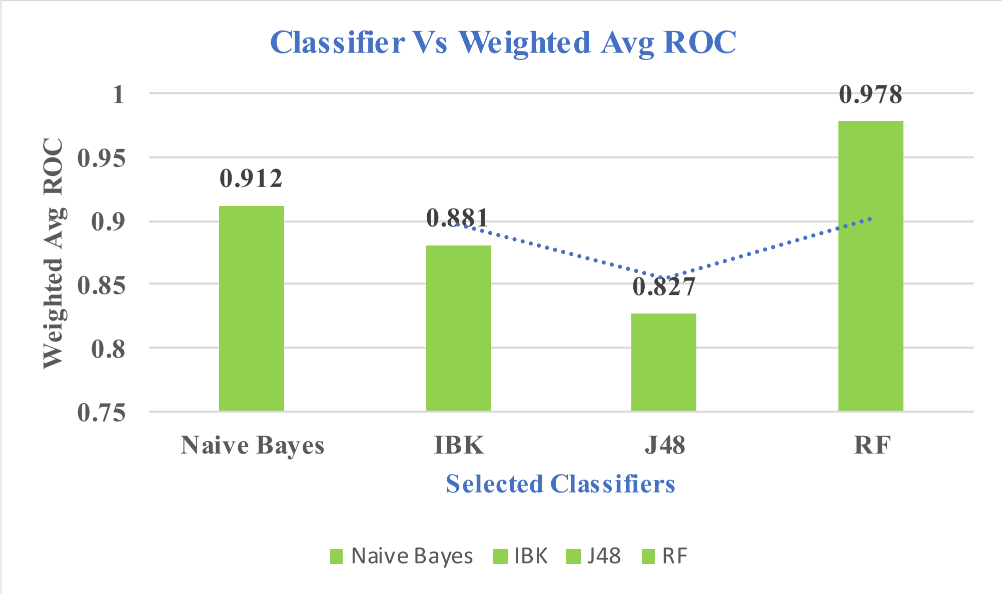

Figure 22 Classifier versus Receiver Operator Characteristics3

V Output Response of Feedback

One of the most powerful algorithms Naïve Bayes algorithm works on the principle of conditional probability and follows a set of Bayes rules, especially for the classification category and produces results of high accuracy and Random Forest that belongs to the tree category whose accuracy is high with less training time. Multiclass object detection is done in RF. Tree-based classifiers, convenient interactive graphical user interfaces are provided for data exploration, for setting up largescale experiments on distributed computing platforms, and for designing configurations for streamed data processing. Classification and prediction algorithm in Weka [22].

To achieve the optimum model a preprocessing tool in weka is used with 631 attributes and 58 instances. The classifying operations are performed using some of the base classifiers. They are Instance-based classifiers, Rules-based classifiers, Decision-based classifiers, and Bayes category. Recent advancement in the Weka tool called SMOTE is added to improve the performance in such a way that we get an optimum model in the feedback loop.

VI Conclusion and Future Scopes

Currently, this article work is a start-up in predicting the stem position of a control valve using image processing techniques and machine learning algorithms. The combination of the two has helped in getting an optimum model in the feedback loop of a second-order system and also help in decreasing the problems related to control valve positioning issues. The accuracy of the system is 92.405% by a selected classifier random forest and 0.978 is the receiver operator characteristics.

The future possible improvements in future work can extend the usability in measuring other parameters like level, flow, and Pressure.

Acknowledgment

I would like to express special thanks of gratitude to our esteemed management of Bharath Institute of Higher Education and Research for encouraging me in writing and exploring research articles.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Serrano, J., Ortiz-Soto, E., Chen, S., Chien, L. et al., "An Efficient Machine Learning Algorithm for Valve Fault Detection," SAE Technical Paper 2022-01-0163, 2022, https://doi.org/10.4271/2022-01-0163.
[2] Liu, H., Zhao, Y., Zaporowska, A. et al. A machine learning-based clustering approach to diagnose multi-component degradation of aircraft fuel systems. Neural Comput & Applic (2021). https://doi.org/10.1007/s00521-021-06531-4[2]
[3] Lin Y, Zakwan S, Jennions I (2020) A Bayesian approach to fault identification in the presence of multi-component degradation. Int J Progn Health Manag 8(1). https://doi.org/10.36001/ijphm.2017.v8i1.2530
[4] Yu J, Zhang C, Wang S (2021) Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis of industrial processes. Neural Comput Appl 33:3085–3104
[5] Subramanian N, He H, Jennions I (2020) Fault detection for aircraft fuel system with neural network. SSRN Electron J. https://doi.org/10.2139/ssrn.3718044
[6] Baudot P, Tapia M, Bennequin D, Goaillard JM (2019) Topological information data analysis. Entropy 21(9):869
[7] Dangut, M.D., Jennions, I.K., King, S. et al. A rare failure detection model for aircraft predictive maintenance using a deep hybrid learning approach. Neural Comput & Applic (2022). https://doi.org/10.1007/s00521-022-07167-8

[8] Farzad A, Gulliver TA. (2019) Log message anomaly detection and classification using auto-B/LSTM and auto-GRU, pp. 1–28
[9] Livieris IE, Pintelas E, Pintelas P (2020) A CNN–LSTM model for gold price time-series forecasting. Neural Comput Appl 32:17351–17360. https://doi.org/10.1007/s00521-020-04867-x
[10] Debayle J, Hatami N, Gavet Y. (2018) Classification of time-series images using deep convolutional neural networks, 23 doi: https://doi.org/10.1117/12.2309486
[11] Lu W, Li J, Wang J, Qin L (2020) A CNN-BiLSTM-AM method for stock price prediction. Neural Comput Appl 33:4741–4753. https://doi.org/10.1007/s00521-020-05532-z
[12] Zhao B, Lu H, Chen S, Liu J, Wu D (2017) Convolutional neural networks for time series classification. J Syst Eng Electron 28:162–9. https://doi.org/10.21629/JSEE.2017.01.18
[13] Ouhame S, Hadi Y, Ullah A (2021) An efficient forecasting approach for resource utilisation in cloud data center using CNN-LSTM model. Neural Comput Appl 33:10043–10055. https://doi.org/10.1007/s00521-021-05770-9
[14] Munna MTA, Alam MM, Allayear SM, Sarker K, Ara SJF (2020) Prediction model for prevalence of type-2 diabetes complications with ANN approach combining with K-fold cross validation and K-means clustering, vol 69. Springer, Berlin
[15] David Dangut M, Skaf Z, Jennions I. (2020) Rescaled-LSTM for predicting aircraft component replacement under imbalanced dataset constraint. In: 2020 Adv. Sci. Eng. Technol. Int. Conf. ASET 2020, doi: https://doi.org/10.1109/ASET48392.2020.9118253
[16] Kamath U, Liu J, Whitaker J (2019). Deep Learning for NLP and Speech Recognition. https://doi.org/10.1007/978-3-030-14596-5
[17] Lawrence I. Oborkhale,Okoro C. K.,Omosun Yerima,ChidiebereE. Onuoha,Prediction of Control Valve Cavitation Using Machine Learning, Technique,IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 15, Issue 5 Ser. I (Sep. – Oct. 2020), PP 26-35, www.iosrjournals.org
[18] K. S. Pawar, S. Sondkar, S. Dattarajan and N. Fernandes, "Comparative analysis of fuzzy logic and machine learning algorithm for predictive analysis of control valve," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), 2019, pp. 1-6, doi: 10.1109/ICCUBEA47591.2019.9128809.
[19] Sheena Angra and Sachin Ahuja, "Machine learning and its applications: A Review", 2017 International Conference On Big Data Analytics and computational Intelligence(ICBDACI).
[20] Pariwat Ongsulee, Veena Chotchaung, Eak Bamrungsi and Thanaporn Rodcheewit, "Big Data Predictive Analytics and Machine Learning", 2018 Sixteenth International Conference on ICT and Knowledge Engineering.
[21] Parth Wazurkar, Robin Singh Bhadoria and Dhananjai Bajpai, "Predictive Analytics in Data Science for Business Intelligence Solutions", 2017 7th International Conference on Communication Systems and Network Technologies.
[22] Aljubran, Mohammad, Horne, Roland, Prediction of Multilateral Inflow Control Valve Flow Performance Using Machine Learning, VL - 35, SPE Production & Operations, 2020/04/01, DOI:10.2118/196003-PA
[23] WEKA: http://www.cs.waikato.ac.nz.

Authors Profile

Mrs B. Kalaiselvi working as Assistant Professor and Part Time Research Scholar in the department of Electronics and Communication Engineering, Bharath Institute of Higher education and Research, Chennai. She Received her M. Tech degree in Electronics and Instrumentation Engineering form Bharath University, Chennai, in 2014. She has received her B.E Electronics and Communication Engineering from Satyabhama University, Chennai in 2011. Her current area of interest is Measurements and Instrumentation, Digital image Processing Artificial Intelligence and Machine Learning. Email:kalaigopal1973@gmail.com

Dr. B. Karthik is an Associate professor in the Department of Electronics and Communication Engineering, Bharath Institute of Higher Education and Research, Chennai, India. He received his PhD (ECE) – Image Processing in May 2017. He received his Master of Engineering - Applied Electronics at Thiruvalluvar College of Engineering & Technology in 2011. He received his Bachelor of Technology in Electronics and Communication Engineering at Bharath Institute of Higher Education and Research, Chennai, India in 2007. His area of interests includes Image & Video Processing, Cryptography, and Network Security System Techniques. He is a member of IEEE, ITEEA, IAIP, CSTA and IAENG. Email: :karthikguru33@gmail.com

Dr A. Kumaravel is working as a Professor and Dean, School of Computing, Bharath Institute of Higher Education and Research, Chennai. His research interest includes Functional Paradigm, Soft Computing, Cloud Computing, Machine Learning, Computational Intelligence and Knowledge Engineering. He is a life Member of ISTE and IET. He is a reviewer for IEEE transactions and Elsevier journals. Email:drkumaravel@gmail.com