AutoPred: A Fault Diagnosis System for Boilers using AutoML

Akanksha Singh Pandaw, Aparna Sinha *Student Member, IEEE* and Debanjan Das *Senior Member, IEEE*

***Abstract*—Boilers are one of the most vital assets for the production of energy in thermal power plants. They require routine maintenance and inspection because of their sizeable investment and severe working conditions. A predictive health analysis of such equipment is crucial to increase their efficiency and prevent any unforeseen mishap. The existing technologies for fault detection are either based on condition monitoring or personnel expertise. Since boilers are operated in a harsh working environment, there is only a limited time to act upon the faults that has occurred. The Machine Learning (ML) models have the ability to analyze and predict a vast set of data in a short span of time, which makes them a dependable assistance for diagnosing faults in industrial equipments. Thus, there is a need to develop a reliable and robust system that could predict the working conditions and provide necessary alerts about the faults on time. This study proposes a Predictive Health Maintenance (PHM) system that predicts the upcoming faults and generate alert messages with their root cause analysis. A comparative study between various ML algorithms shows that Automated Machine Learning (Auto ML) gave predictions with highest accuracy of 98.32% by choosing the random forest as the classifier. The root cause analysis provided with the alert messages saves time and resources for rectifying the faults and taking preventive measures. Hence, fault diagnosis using machine learning algorithms can help increase the boiler's lifespan and effectiveness.**

***Index Terms*—Thermal Power Plants, Predictive Health Maintenance, Machine Learning, Random Forest, Auto ML**

I. Introduction

THE Thermal Power Plants (TPP) are one of the primary resources for the generation of electricity. Boilers are widely used in TPP for several industrial productions [1]. These boilers are made up of complex systems having interconnection of various sub-units to perform a specific task [2]. Failure of any subsystem affects the performance of the whole boiler. In recent years, the electricity demand has grown exponentially. Thus, it is crucial to maintain the boiler continuously in the operating state to meet the growing needs. Its harsh working environment creates a situation of inevitable equipment failure, high accident rates, and economic losses [3]. However, it could be minimized by following a proper maintenance strategy. Due to its system complexity, relying just on operating personnel is not recommended because conducting a comprehensive understanding, determining the root cause quickly, and generating an alert in a short time is extremely difficult [4]. Therefore, there is a need to adapt such technology that could diagnose multiple failures effectively in a short span of time.

With the emergence of Industry 4.0, autonomous systems, machine learning, and predictive maintenance are being used extensively in various industries for conserving industrial equipment in the longer run [5]. The combination of these approaches is useful for handling unseen equipment failures and resource management, thus saving both time and cost. It is important to prognosis any abnormal conditions in advance for ensuring safety, reliability, and availability of various industrial processes such as chemical factories, automobiles, and power plants. The faults in industrial equipment can cause severe threats to life and materials [6]. Therefore it is crucial to use a fault diagnosis system that could detect the abnormal operating conditions by analyzing complex and non-stationary behaviors of process parameters and help assist in executing proper actions at the initiatory stage of faults. Machine learning techniques are an important factor in forecasting faults and their root cause analysis [7]. If an inappropriate technique is selected, it may cause time loss and infeasible maintenance scheduling.

Over the years, various fault diagnosis techniques for boilers in thermal power plant has been developed for sustainable maintenance of the system. In a study by Olsson *et al.* [8], operational data was collected from actual installations independent of a reference system. For both the forecasting and estimation of the degradation, the linear regression approach was applied, and the results obtained showed a similarity of performance for both the models. This stresses the importance of choosing a correct machine learning model for predictive maintenance. Boilers are a critical unit of a co-generation system. The operative parameters in the boiler‘s burner system are measured and characterized to obtain a specific set of descriptors. [9] Kim *et al.* used the data mining approach for analyzing these sets effectively. The knowledge of an equipment performance extracted beforehand gives leverage for generating an intelligent alarm system. [10] developed novel fault detection and identification method for plugged tubes using the Principal Component Analysis (PCA). The temperature was taken as the parameter to identify the characteristics of blocked tubes.

While condition-based monitoring reduces downtime, it heavily relies on real-time sensor measurements. Keeping equipment well-maintained requires investment and more labor-intensive work [11]. It also does not assure to provide enough time for acting upon the faults. While many researchers have proposed methods to diagnose various types of faults in industrial equipment, the occurrence of hybrid

Akanksha Singh Pandaw is with the Department of Computer Science & Engineering, IIIT Naya Raipur, Email: akanksha19100@iiitnr.edu.in.

Aparna Sinha is with the Department of Electronics & Communication Engineering, IIIT Naya Raipur, Email: aparna@iiitnr.edu.in.

Debanjan Das is with the Department of Electronics & Communication Engineering, IIIT Naya Raipur, Email: debanjan@iiitnr.edu.in.

faults is yet to be considered [12]. The inception of auto-machine learning has opened up the possibilities of instinctively selecting the most appropriate algorithm with hyper-tuned parameters, thus having the potential to be effectively used in the area of predictive maintenance [13].

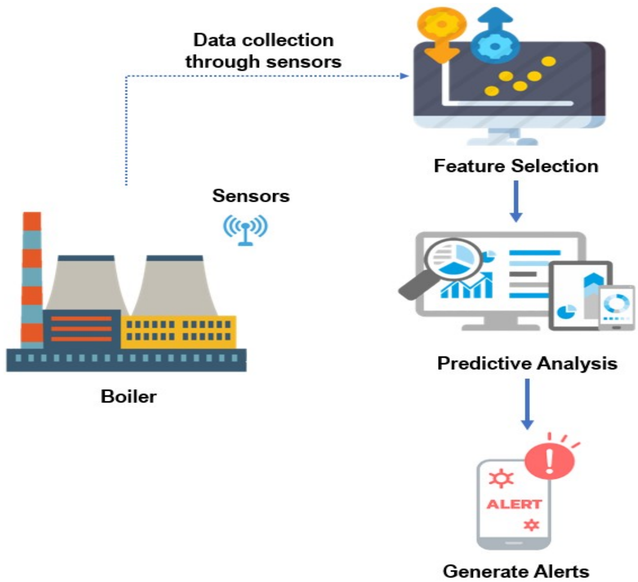

Fig. 1. Overview of the proposed system

The main objective of this study is to develop a predictive health maintenance system that could analyze the root cause and predict the occurrence of faults in the boiler of thermal power plants. Figure 1 describes an overview of the proposed solution. The process starts with collecting data from the boiler using the embedded sensors. The data collected directly from the boiler contains several features which may or may not impact the occurrence of faults. Therefore, feature selection is applied on the raw data to train the ML model only on useful features which in turn, also reduces its computational complexity. This is followed by predictive analysis of the faults using ML models based on the data received. This step will help in taking appropriate measures to optimize the useful life of the boiler. After the predictions, an alert message is generated that informs about the occurrence of impending fault, along with its root cause analysis. This system is crucial for taking precautionary measures to rectify the faults ahead of time.

II. Related Works

A. Related Prior Research Works

Faults and failures in any equipment are inevitable. It can lead to catastrophic problems on both human and economic levels. The fault diagnosis solutions highly depend on the sensor network and connected objects [21]. Therefore, there is a need to design, develop, and fabricate different types of sensors and sensing technology that could accurately locate unforeseen defects in any structure in an offline and online manner. Tang *et al.* [22] addressed the limitations of traditional fault diagnosis methods and proposed a convolutional neural network (CNN) based fault diagnosis method in rotating machinery. However, the technique could not produce a desirable classification performance due to the absence of feature selection. Another work by Shao *et al.* [23] explored the deep learning framework, which uses transfer learning for diagnosing the faults in motors, gear-boxes, and bearings.

Boilers in TPP have a significant contribution to the generation of power. Therefore, their fault diagnosis is of paramount importance for their long-run performance. Azari textitet al. [20] adopted a combination of data-driven methods using a neuro-fuzzy inference system algorithm to improve the robustness of fault diagnosis methods in boilers. In reality, two or more faults can occur, which is still needed to be explored in such systems. In an another approach, the normal conditions were compared with the abnormal ones to analyze the fault detection system in boilers [24]. But without any feature selection, the processing of real data is not desirable due to various irrelevant features. In a study by Zhang *et al.* [26], the advantages of deep learning (DL) algorithms over the machine learning (ML) methods have been studied to carry out the diagnosis of bearing faults. Another study by Liu *et al.* [18] discusses the time-consuming behavior of DL models and accordingly proposed a CNN and Transfer Learning (TL) method for developing a fault diagnosis system. Despite that, the methods were based on supervised learning, and therefore, it will be prone to misclassification on the unlabeled dataset. A guiding system for the operation of boilers in thermal power plants has been developed by Cui *et al.* [27] using the Principal Component Analysis (PCA). It was concluded from the results that the model was successful in finding the influencing factors of the boiler's temperature deviation. However, when the relationship between the two factors is weak, PCA does not perform well to reduce the data. Li *et al.* [28] proposed a Convolutional Neural Network (CNN) framework to extract the complicated features from a time-series dataset of boiler's vibration analysis. The method encodes one-dimension time series into two-dimension images. Although, the characteristic learning ability of such networks is limited due to their shallow structure.

Over the years, a large number of predictive maintenance strategies have been developed and applied to industrial systems. In a study by Chen *et al.*, [33] a dynamic predictive maintenance scheduling using deep learning is developed for the health prognosis of the industrial systems. Applying this strategy to a combination of components and failures is still demanding. Zhang *et al.* [12] presented a comprehensive study which focused on data-driven methods for predictive maintenance (PdM). However, the data training has been done using supervised learning, which will not be useful when the learning is carried out on the unlabeled data. Machine Learning (ML) technology has also contributed significantly to PdM over the past few years. A study by Soleh *et al.* [30] utilized the ML technology for anomaly detection based on regression models for time series data. The rapid development of artificial intelligence and industrial Internet of Things (IIoT) technologies have received considerable attention from researchers to develop a new technology, intelligent predictive maintenance (IPdM). This has been explored in a study by

TABLE I
COMPARISON BETWEEN EXISTING TECHNOLOGIES FOR FAULT DIAGNOSIS

Method Used	**Type of Data**	**Data Distribution**	**Feature Ranking**	**Predictive**	**Alert Generation**	**Accuracy**
Acoustic Steam Leak [14]	Real-time	Balanced	×	×	×	95.23%
k-means clustering [19]	Simulated	Imbalanced	×	×	×	99.2%
Support Vector Machine [15]	Real-time	Balanced	×	✓	×	88.2%
Regression [16]	Real-time	Imbalanced	×	✓	×	88.18%
Markov probabilistic model [17]	Real-time	Balanced	×	✓	×	99.98%
Transfer CNN [18]	Real-time	Balanced	✓	×	×	98.7%
Decision Tree [12]	Simulated	Imbalanced	✓	✓	×	97.8%
Minimum Spanning Trees - k means [20]	Simulated	Imbalanced	×	✓	×	*
AutoPred	Simulated	Imbalanced	✓	✓	✓	98.32%

* Not available

Zhang *et al.* [31] for predicting future signal samples using an enhanced attention mechanism. Although, there is always a risk of faults in sensing devices like in IoT. Huynh textitet al. [32] worked in a parametric predictive maintenance decision frame which leverages the information of system remnant life in maintenance decisions. However validation of the framework with real data is missing. Boilers in TPP have a significant contribution to the generation of power. Therefore, their fault diagnosis is of paramount importance for their long-run performance. Azari *et al.* [20] adopted a combination of data-driven methods using a neuro-fuzzy inference system algorithm to improve the robustness of fault diagnosis methods in boilers. In reality, two or more faults can occur, which is still needed to be explored in such systems. In another method, the normal conditions were compared with the abnormal ones to analyze the fault detection system in boilers. [24]. But without any feature selection, the processing of real data is not desirable due to various irrelevant features.

A comparision between existing technologies for fault diagnosis of boilers is showed in Table I.

B. Issues with the existing solutions

- The current scenario for diagnosing a fault in industrial equipment relies on personnel expertise or condition-based monitoring. These approaches are highly subjective and do not give enough time to act upon the faults occurred.
- Root cause analysis of faults is not considered along with identification of faults, even though it is a crucial parameter to prevent any unforeseen conditions.
- Prediction of faults using machine learning algorithms can become a computationally costly process if a vast amount of input data is received from a real-time operating machines. An approach which can reduce this complexity is needed to be applied on the raw data, in order to achieve a better model performance.

C. The vision of AutoPred

AutoPred is envisioned to be a predictive health maintenance system for boilers in thermal power plants, that could predict any unforeseen conditions and alert the workers to prevent equipment failure.

D. Proposed Solution

- AutoPred is a predictive health maintenance system for boilers in thermal power plants, that predict faults and generate alerts using machine learning algorithms on a simulated dataset.
- The system aims to intercept the occurrence of faults ahead of time by analyzing its primary root causes and the level of component degradation due to the fault.
- A rapid and computationally low-cost approach using the feature selection method is applied on the raw dataset. This enabled the ML model to analyze only those variables that impact the occurrence of faults in the boiler.

E. Novel Contributions

- The proposed system uses AutoML for the prediction of faults. Its usage has enabled the selection of the most appropriate model with optimized hyperparameters that resulted in higher efficiency of the system while still retaining the model quality.

- The feature ranking is established according to their impact on faults, and dimensionality reduction of the dataset is implemented accordingly. This approach resulted in decreased training time of the ML model and an increased accuracy.
- The alert messages are generated that identifies the fault and its root cause analysis, describing its source and the percentage of the component degradation.

III. Methodology

A. Dataset Collection

In this study, the data used for predicting the faults is proposed by [12]. The samples for the dataset were developed and validated on a boiler emulator capable of simulating regular operation and operation under critical fault conditions. The emulator represents the Viessmann Vitorond 200 Gas Fired Boiler VD2 Series, 380 model. The validation of this model was done by replicating ANSI/AHRI Standard 1500 – Performance Rating of Commercial Space Heating Boiler. The training and test data were created by simulating potential fault conditions and varying model parameters, either individually or in combination. The resulting input and output data are labeled with the fault name. The simulation yielded imbalanced distribution for five categories of faults, 'Lean', 'Nominal', 'Excess Air', 'Fouling', and 'Scaling'. Figure 2 shows the distribution of all the 5 classes of faults. To allow the progression of a fault condition to be detected over time, the complete set of conditions 31 classes was used for the initial classification, all with equal number of samples.

The class 'Nominal' refers to the normal working conditions of the boiler. In this state, the performance of the boiler is optimal and all the preventive measures are taken to keep the boiler at this nominal condition. In an operating boiler, 'Excess Air' represents the heat loss, caused due to the presence of extra amount of air that is needed for combustion. Likewise, 'Lean' means that sufficient amount of air is not present and thus steam is not produced at needed capacity. 'Scaling' in boilers refers to the formation of deposits on the inside of piping due to the precipitation of impurities by heat surfaces [33]. These deposits hinder the process of heat transfer and cause system failures. In 'Fouling', the particles cling to the convection heat surfaces such as superheater and reheaters, but can be removed relatively easily by mechanical means such as hydro blasting or scrubbing with a soft bristle brush [33].

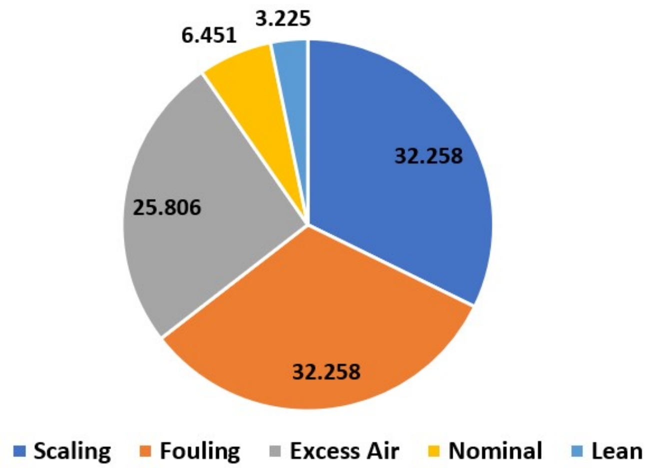

Fig. 2. Distribution of samples for five boiler faults (in percentage)

B. Fault Analysis

For the predictiion and classification of fault classes, a comparative study was carried out between five different ML algorithms; Decision Tree, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), XG Boost, and Random Forest. The dataset was split into train set (70%) and test set (30%). 'Condition' and 'Class' were the two co-dependent target variables present in dataset, out of which the 'Condition' variable was selected as the target class due to its more descriptive nature than the 'Class' variable. The initial training was done without applying the feature selection to the dataset. After the model training, it was found out that Decision Tree had the highest accuracy (95.37%), and KNN had the lowest accuracy amongst all (54.86%). The evaluation also considered the inter-class accuracies, calculated using the Equation (1) where TP, TN, FP and FN represents True Positive, True Negative, False Postive and False Negative class respectively. Satisfactorily results for five different ML models as shown in Table II.

TABLE II
Inter-class accuracies for different machine learning models

Algorithm	Lean	Nominal	Excess Air	Scaling	Fouling
Decision Tree	0.98	0.91	0.95	0.95	0.99
Random Forest	0.98	0.86	0.96	0.87	0.96
KNN	0.98	0.81	0.95	0.81	0.93
SVM	0.99	0.81	0.95	0.84	0.97
XG Boost	0.99	0.89	0.95	0.83	0.95
Auto ML	0.99	0.96	0.97	0.94	0.99

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The training of five different algorithms to evaluate and analyze their performance came out to be time taking and computationally costly. In such a scenario, an approach that can give high performance with less complexity was required to meet the current objectives.

Automated Machine Learning (Auto ML) is a significant trend in data science which automates the process of models' selection, composition, and parameterization for the given dataset. This approach eased up the pipeline creation and hyperparameter tuning, selecting the most appropriate ML model with the best-suited hyper-parameters and increased accuracy [34]. Figure 3 [35] explains the working of automated machine learning. The user gives three inputs to the system of auto ML. First is the dataset, upon which the ML algorithms are applied on. Second is the target metric, which is the variable that the system will predict and lastly, the time constraints. This input is crucial because, it decides the total amount of time that will be dedicated for the training of each individual model in the Auto ML. After each model is trained, their accuracy scores are compared and the model with the highest accuracy is chosen to be the final model for training the data.

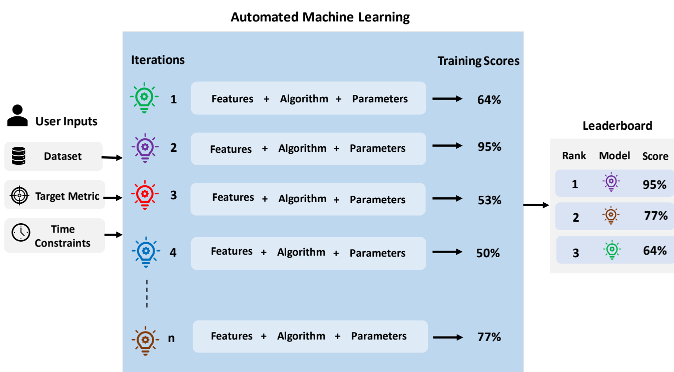

Fig. 3. Working of Automated Machine Learning

In this work, the total training time for the Auto ML was set to 90 minutes and around 120 seconds were given for the individual training of each model. After the training, the Auto ML selected the Random Forest as the classifier with an accuracy of 98.32%. The difference between the accuracies of Random Forest with and without AutoML arises due to the fact that, with the former case, the model is hyper-tuned and optimized during the training automatically. Thus, this approach enhances the overall performance of the algorithm.

C. Feature Selection

To optimize the performance further, feature selection is applied to the dataset and trained again with AutoML and Decision Tree (since they were the top 2 performing model. The impact of all the feature variables was observed on all the five classes of faults, and it was concluded that the variable 'Treturn' has little to no impact on the target class. Therefore, dimensionality reduction was carried out by removing the 'Treturn' variable from the dataset, and once again the model is trained, but this time with only the impactful variables. The results after the model training showed a considerable increase in accuracy in both AutoML and Decision Tree, as shown in Figure 9. This method also decreases the classifier's complexity due to a lesser number of training variables.

Algorithm 1 illustrates overall approach used for predictive analysis of boilers using the Auto ML. All the features in the dataset are ranked according to their impact on the occurence of faults. The features with lesser or no impact are removed and a new dataset is formed for training the model. Using the Auto ML, most appropriate classifer is selected for the data, which here is, Random Forest (RF). The RF starts the training process by creating decision trees from the data samples using the Gini Impurity Measure as shown in Equation (2).

$$\sum_{i=1}^{N} P_{i,K}(1 - P_{i,K}) = 1 - \sum_{i=1}^{N} P_{i,K}^2 \tag{2}$$

In the above equation, N is the list of classes. In this case N= 'Nominal', 'Lean', 'Excess Air', 'Scaling', 'Fouling'. K represents category and $P_{i,K}$ is the probability of category K having the class i.

Thereafter, each decision tree makes a prediction for the fault class. In the last step, voting takes places between all the predictions and the most votes one is selected to be the final prediction result.

Algorithm 1 : *Predictive Analysis using Random Forest Classifier*

Input: Data collected from the sensors
Output: Alert Message with Root Cause Analysis

1: Feature Selection extracts out the all the features with significant impact on target variables
2: The Auto ML starts the model training with the newly selected dataset and choose RF as the classifier
3: RF takes the original dataset and create N bagged samples of size n, with n smaller than the original dataset
4: Construct decision tree for each N samples by randomly selecting M features, from all the features in training set using Gini impurity measure
5: Prediction is generated from every decision tree
6: Most Voted Prediction among all decision trees is the final prediction result

A visual representation of the working of Random Forest is shown in Figure 4.

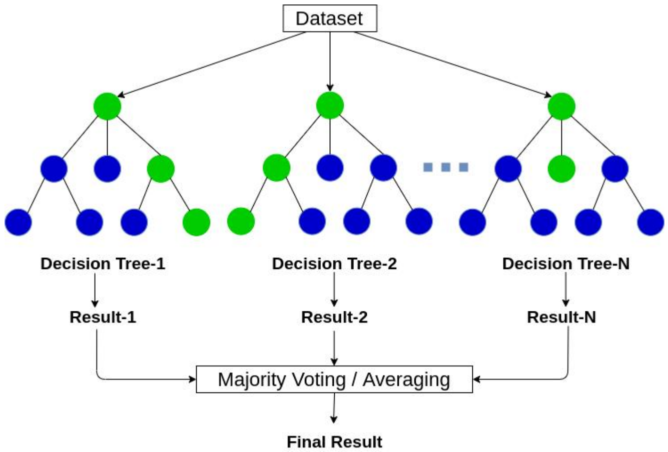

Fig. 4. Working of Random Forest for the prediction of fault class

D. Root Cause Analysis

The target variable ‘Condition’ consists of 31 unique fault conditions that describe the state of boiler components after the fault has occurred. These descriptions set the basis for root cause analysis. An alert message is generated after the fault prediction that has descriptions about the working state of the components. The alert will help analyze the primary cause of the faults beforehand and take preventive measures accordingly.

IV. Results

A. Dataset

Figure 5 illustrates the correlation matrix between five feature variables. From the figure, it can be observed that; there is a negative correlation between ‘Treturn’ - ‘Tsupply’ and ‘Water_Mdot’ - ‘Tsupply’. This indicates that the variables are inversely proportional to each other. Also, there is a positive correlation between ‘Tsupply’-‘Fuel_Mdot’ and ‘Tsupply’-‘Tair’ which means that the variables are proportional to each other. ‘Water_Mdot’-‘Fuel_Mdot’, ‘Water_Mdot’-‘Tair’, ‘Tair’-‘Fuel_Mdot’, ‘Treturn’-‘Fuel_Mdot’ and ‘Treturn‘-‘Tair’ showed no correlation between them. These variables are independent of each other and has no predictive relationship.

	Fuel_Mdot	Tair	Treturn	Tsupply	Water_Mdot
Fuel_Mdot	1	0	0	0.0012	0
Tair	0	1	0	0.004	0
Treturn	0	0	1	-0.62	0.96
Tsupply	0.0012	0.004	-0.62	1	-0.66
Water_Mdot	0	0	0.96	-0.66	1

Fig. 5. Correlation matrix for the target variable ‘Class‘ having five unique classes of faults

B. Predictive Analysis

Figure 6 shows the comparision between accuracies for different classifiers. These accuracies were obtained before

	Lean	Nominal	ExcessAir	Fouling	Scaling
Lean	0.99	0.0088	0	0	0
Nominal	0.0027	0.93	0.049	0	0.016
ExcessAir	0	0.33	0.67	0	0
Fouling	0	0	0	0.89	0.11
Scaling	0	0.015	0	0.023	0.96

Fig. 7. Comparision of accuracies before and after applying the feature selection

applying feature selection to the dataset. As shown in the figure, Auto ML performs with the highest accuracy while KNN (with k=3) had the lowest performance. These accuracies were obtained before applying feature selection to the dataset.

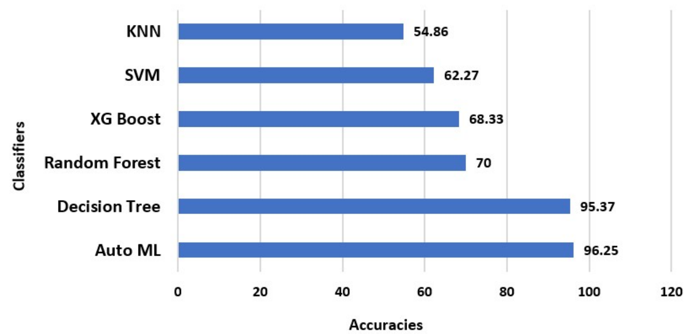

Fig. 6. Comparision of accuracies between different ML models

Figure 7 illustrates the confusion matrix between the 5 classes of faults. False positive and false negative values for all the faults were found to be significantly low as compared to the true positive and true negative values.

C. Feature Selection

The feature selection not only reduces the training time of the ML models, but it is also considers the impact of each feature on the boiler faults. Figure 8 shows the effect of 5 different feature on the faults. The ‘Treturn’ feature had no impact on any of the fault events. Thus, it is removed from the dataset and the classifiers are trained again with only the impactful features in the dataset.

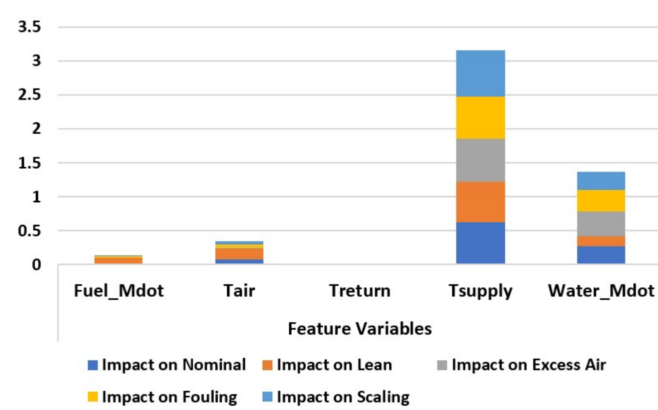

Fig. 8. Impact of features on different classes of faults

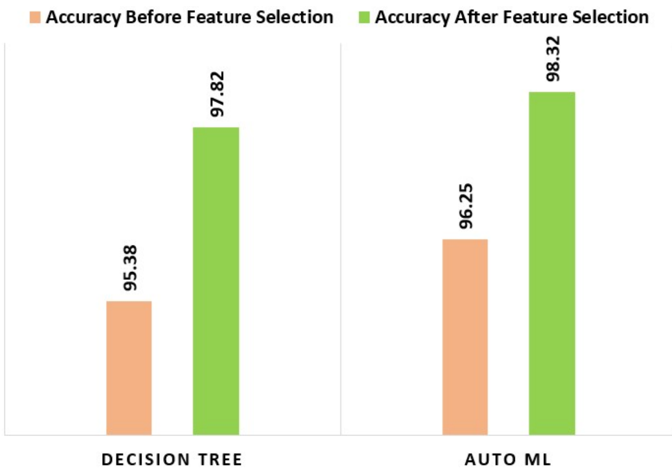

Fig. 9. Comparision of accuracies before and after applying the feature selection

D. Impact of Feature Ranking on model performance

Figure 9 depicts the performance of Decision tree and Auto ML after applying the feature selection to the dataset. The increase in the accuracy shows the importance of feature ranking in predictive models.

E. Generation of Alert Messages

In Table III, different scenarios has been shown where the alert messages are generated. The message briefly describes the name of the fault, name of its root cause and level of degradation caused due that fault. The incoming parameteres can be for real-time data or the future ones. For real-time, the message will alert about the fault that is currently in the boiler and for the upcoming data, the messages will alert about the faults that is going to be occurred in the boiler. This will help in looking upon the root causes of faults and take preventive measures ahead of time.

TABLE III
ALERT MESSAGES GENERATED FOR DIFFERENT FAULT SCENARIOS

Incoming Parameters	Alert Message
[‘4’, ‘297’, ‘333.0000002’, ‘335.252335’, ‘9’]	Boiler has 46.0% of Scaling
[‘1’, ‘283’, ‘333.0000002’, ‘342.3244007’, ‘9.5’]	Boiler operating at Nominal condition
[‘1’, ‘283’, ‘333’, ‘356.3152528’, ‘3.5’]	Boiler opearting at 20.0% Excess Air
[‘1’, ‘283’, ‘333’, ‘359.2389633’, ‘3.5’]	Boiler opearing at 5.0% Lean condition
[‘1’, ‘287’, ‘333.0000001’, ‘337.9953434’, ‘8’]	Boiler has 31.0% of Fouling

V. Conclusion and Future Work

Predictive Health Maintenance is an important aspect for the industrial equipments. Knowing the system failures and hazards beforehand gives the leverage of time to act upon the faults caused and save various worthy resources. In this study, a simulated dataset formed using a boiler emulator was used to predict faults in the boiler of thermal power plants. In a comparative study between five different classifiers it was observed that, decision tree outperforms the KNN, SVM, Random Forest and XG Boost classifiers. To further increase the performance, feature ranking was used to reduce the computational cost and the accuracy was recorded as 98.32%. The prediction of the fault is followed by providing alert messages with descriptions about its root cause. This alerts will help in taking proper actions in-advance to prevent the system failures.

For future work, hybrid faults will also be investigated and considered for the predictive analysis. The proposed system has a potential to be deployed in real time using sensors and hardwares, which will also be explored in the future works.

References

[1] K. Xiangyu, Y. Yang, Z. Meng and Y. Hongzhi, “Design and Simulation of Boiler Combustion Control System in Thermal Power Plant,” 2021 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 83-87, 2021

[2] Patel, Himanshukumar R. and Dr. Vipul A. Shah. “Fault Detection and diagnosis methods in power geneartion plants- The Indian power geneartion sector perspective: an introductory review 3”, Semantic Scholar, 2018.

[3] Swiercz, M., “Mroczkowska, H. Multiway PCA for Early Leak Detection in a Pipeline System of a Steam Boiler—Selected Case Studies”, Sensors, 2020.

[4] S. Thota, R. P. Mandi and S. Chaudhari, ”Survey on Intelligent Control Approaches for Prediction of Boiler Efficiency in Thermal Power Plant,” 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 79-83, 2018

[5] Cınar, Z.M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., Safaei B., “Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. Sustainability”, 2020.

[6] N. G. Lo, J. Flaus and O. Adrot, ”Review of Machine Learning Approaches In Fault Diagnosis applied to IoT Systems,” 2019 International Conference on Control, Automation and Diagnosis (ICCAD), pp. 1-6, 2019

[7] H. A. Tokel, R. A. Halaseh, G. Alirezaei and R. Mathar, “A new approach for machine learning-based fault detection and classification in power systems,” 2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1-5, 2018

[8] Tomas Olsson, Enislay Ramentol, Moksadur Rahman, Mark Oostveen, Konstantinos Kyprianidis, “A data-driven approach for predicting long-term degradation of a fleet of micro gas turbines”, Energy and AI, Volume 4, 100064, ISSN 2666-5468, 2021.

[9] K. h. Kim, H. seok Lee, J. h. Kim and J. Ho Park, “Detection of Boiler Tube Leakage Fault in a Thermal Power Plant Using Machine Learning Based Data Mining Technique,” 2019 IEEE International Conference on Industrial Technology (ICIT), pp. 1006-1010, 2019

[10] Jungwon Yu, Jaeyeong Yoo, Jaeyel Jang, June Ho Park, and Sungshin Kim. “A novel plugged tube detection and identification approach for final super heater in thermal power plant using principal component analysis” Energy, vol. 126, 2017.

[11] F. Yüce and M. Hiller, “Condition Monitoring of Power Electronic Systems through Data Analysis of Measurement Signals and Control Output Variables,” in IEEE Journal of Emerging and Selected Topics in Power Electronics, 2021

[12] W. Zhang, D. Yang and H. Wang, “Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey,” in IEEE Systems Journal, vol. 13, no. 3, pp. 2213-2227, Sept. 2019

[13] K. Chauhan et al., ”Automated Machine Learning: The New Wave of Machine Learning,” 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), pp. 205-212, 2020

[14] P. Anujaa, R. Babysree, S. Divya Prabha, D. Karthika, R. Selvaraj, “Acoustic Steam Leak Detection System”, International Jounral of Engineering Research Technology (IJERT) ETCAN – 2020 (Volume 8 – Issue 09), 2020

[15] K. h. Kim, H. seok Lee, J. h. Kim and J. Ho Park, “Detection of Boiler Tube Leakage Fault in a Thermal Power Plant Using Machine Learning Based Data Mining Technique,” 2019 IEEE International Conference on Industrial Technology (ICIT), pp. 1006-1010, 2019

[16] Marek Moleda, Alina Momot, Dariusz Mrozek, “Predictive Maintenance of Boiler Feed Water Pumps Using SCADA Data”, Sensors, 2020
[17] Jagtap, H. Bewoor, A. Kumar, R. Ahmadi, M. Hossein, Lorenzini, Giulio, “ Markov-based performance evaluation and availability optimization of the boiler-furnace system in coal-fired thermal power plant using PSO”, Energy Reports, vol. 6, pp. 1124-1134, 2020.
[18] Q. Liu and C. Huang, “A Fault Diagnosis Method Based on Transfer Convolutional Neural Networks,” in IEEE Access, vol. 7, pp. 171423-171430, 2019
[19] J. Yu, J. Jang, J. Yoo, J. H. Park, and S. Kim, “A Clustering-Based Fault Detection Method for Steam Boiler Tube in Thermal Power Plant,” Journal of Electrical Engineering and Technology, vol. 11, no. 4. The Korean Institute of Electrical Engineers, pp. 848–859, 2016
[20] M. S. Azari, F. Flammini, M. Caporuscio and S. Santini, “Data-Driven Fault Diagnosis of Once-through Benson Boilers,” 2019 4th International Conference on System Reliability and Safety (ICSRS), pp. 345-354, 2019
[21] C. Furse, M. Kafal, R. Razzaghi and Y. -J. Shin, “Special Issue on Embedded Sensors for Fault Diagnosis in Electrical Wiring Interconnection Systems, Power Grids, Structural Cables, Pipelines, and Electrical Machines,” in IEEE Sensors Journal, vol. 21, no. 2, pp. 886-887, 15 Jan,2021
[22] S. Tang, S. Yuan and Y. Zhu, “Convolutional Neural Network in Intelligent Fault Diagnosis Toward Rotatory Machinery,” in IEEE Access, vol. 8, pp. 86510-86519, 2020
[23] S. Shao, S. McAleer, R. Yan and P. Baldi, “Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning,” in IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2446-2455, April 2019
[24] S. Mohod and A. Raut, “PLC SCADA Based Fault Detection System for Steam Boiler In Remote Plant,” 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pp. 1007-1010, 2019
[25] G. Madrigal Espinosa, C. M. Astorga Zaragoza, M. Vazquez Roman, G. L. Osorio Gordillo and M. Adam Medina, “Fault Diagnosis in Sensors of Boiler Following Control of a Thermal Power Plant,” in IEEE Latin America Transactions, vol. 16, no. 6, pp. 1692-1699, June 2018
[26] S. Zhang, S. Zhang, B. Wang and T. G. Habetler, “Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review,” in IEEE Access, vol. 8, pp. 29857-29881, 2020
[27] Y. -j. Cui, L. -w. Xia, Y. Huang and X. -c. Ma, “Research on Fault Diagnosis and Early Warning of Power Plant Boiler Reheater Temperature Deviation Based on Machine Learning Algorithm,” 2020 IEEE 6th International Conference on Control Science and Systems Engineering (ICCSSE), 2020
[28] C. Li, J. Xiong, X. Zhu, Q. Zhang and S. Wang, “Fault Diagnosis Method Based on Encoding Time Series and Convolutional Neural Network,” in IEEE Access, vol. 8, pp. 165232-165246, 2020
[29] C. Chen, Z. H. Zhu, J. Shi, N. Lu and B. Jiang, “Dynamic Predictive Maintenance Scheduling Using Deep Learning Ensemble for System Health Prognostics,” in IEEE Sensors Journal, vol. 21, no. 23, pp. 26878-26891, 1 Dec.1, 2021
[30] M. Soleh, A. Riyadi and A. P. Asnel, “Development Of Predictive Maintenance Methodology Utilizing Machine Learning Technology To Support Plant Health Management,” 2019 International Conference on Technologies and Policies in Electric Power Energy, pp. 1-6, 2019
[31] W. Zhang, D. Yang, Y. Xu, X. Huang, J. Zhang and M. Gidlund, “DeepHealth: A Self-Attention Based Method for Instant Intelligent Predictive Maintenance in Industrial Internet of Things,” in IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5461-5473, Aug. 2021
[32] K. T. Huynh, A. Grall and C. Bérenguer, “A Parametric Predictive Maintenance Decision-Making Framework Considering Improved System Health Prognosis Precision,” in IEEE Transactions on Reliability, vol. 68, no. 1, pp. 375-396, March 2019
[33] “Typical Causes of Slagging and Fouling Problems in Boilers”, 2015, https://www.powermag.com/typical-causes-of-slagging-and-fouling-problems-in-boilers/
[34] Ben Lutkevich, “automated machine learning (AutoML)”, TechTarget, 2020
[35] Microsft Docs, “What is automated machine learning (AutoML)?”, 2021, https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml