International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 273-288
https://doi.org/10.70917/ijcisim-2026-2329
Article

Explainable AI-Driven Predictive Maintenance Framework for Industrial Equipment in Industry 4.0 - Enabled Smart Factories

Khushboo Kachhawa[1], Pratik Jain[2], Harleen Kaur Malhotra[3], Amit Chandak[4], Satya Narayan Suman[5], Pravin Tathod[6]

[1] Department of Computer Science and Engineering, IPS Academy, Institute of Engineering and Science, Indore, Madhya Pradesh, India, khushbookachhawa.ies@ipsacademy.org
[2] Department of Computer Science and Engineering, IPS Academy, Institute of Engineering and Science, Indore, Madhya Pradesh, India, pratikjain@ipsacademy.org
[3] Department of Computer Science and Engineering, Oriental University, Indore, Madhya Pradesh, India, harleenkaur.935@gmail.com
[4] Department of Mechanical Engineering, IPS Academy, Institute of Engineering and Science, Indore, Madhya Pradesh, India, amitchandak@ipsacademy.org
[5] Department of Computer Science and Engineering, IPS Academy, Institute of Engineering and Science, Indore, Madhya Pradesh, India, satyanarayansuman.ies@ipsacademy.org
[6] Department of Fire Technology and Safety Engineering, IPS Academy, Institute of Engineering and Science, Indore, Madhya Pradesh, India, pravintathod@ipsacademy.org

Abstract: Predictive maintenance has now become one of the most important strategies of enhancing the workability of the system as well as cutting down unplanned downtime in the complex industrial systems. GRID deviation is one of the essential measurements of the imbalance and stress factors in a power grid operation, which requires precise predictability and timely interventions. This paper suggests a full machine learning predictive maintenance system to predict grid deviation using high-scale operation records. The framework combines systematic feature engineering, ensemble learning models, and explainable artificial intelligence methods to conquer the issues of non-linear behavior, data imbalance, and rare extreme events. It used a dataset of 267, 324 observations, including the operational, maintenance, generation, and time variables. The exploratory analysis showed that grid deviation is a highly skewed, heavy-tailed, variable, which has some high-magnitude events, which highlights the constraints of the conventional linear modeling methods. Several of the ensemble learning models were developed and tested within a similar experimental setup, namely, Random Forest, Gradient Boosting Regressor, and Light Gradient Boosting Machine (LightGBM). Comparative findings indicate that each of the models has good predictive performance; nevertheless, LightGBM has better predictive ability on outliers, robustness, and efficiency. In order to make it more transparent and practical, model explainability was implemented with the help of Shapley Additive Explanations (SHAP). The explainability analysis also found the generation alignment variables, which include actual and programmed generation, and station-level statistical factors to be the most influential in predicting the grid deviation. Besides giving high prediction accuracy, the proposed framework also gives actionable insights to grid operators and maintenance planners. In general, the research paper evidences the efficiency of combined using advanced machine learning and explainability methods of predictive maintenance in large-scale power systems.

Keywords: Predictive maintenance; Grid deviation prediction; Machine learning; LightGBM; Ensemble learning; Feature engineering; SHAP explainability

1. INTRODUCTION

In factories, both the dependability of equipment and the consistency of the production process are the primary factors of efficiency of operations, cost effectiveness, and competitiveness. The direct and indirect costs of such industrial processes include unplanned equipment failures and related downtime which cost a lot of money in maintenance budgets, production schedules, and supply chain commitments all over the world (Carvalho et al., 2019; Meitz, 2025). The classical approaches to maintenance like reactive maintenance that reacts to the failures at the end of their life and time-based preventive maintenance that takes the life cycle of the component as a given constant usually result in high maintenance expenses, non-optimal resource utilization, and high risk of unplanned downtime (Kane et al., 2023). These shortcomings of the traditional methods have prompted the growing academic focus on more active and data-intensive maintenance paradigms, especially Predictive Maintenance (PdM), which uses very large amounts of sensor and operational data to both anticipate when failures will happen and to make optimal maintenance decisions (Meitz, 2025; IJCRT, 2025). PdM is a radical change in the conventional approach since it allows measuring the equipment condition in real time and predicting the possibility of breakdown with the assistance of sophisticated analytics.

Combining these technologies has not only enhanced the accuracy of equipment health monitoring, but also demonstrated a greater range of analytical PdM systems than simple threshold-based alarms to forecasting-based models that approximate Remaining Useful Life (RUL) of critical components (Carvalho et al., 2019; Ramzan, 2025). Due to its capacity to estimate complicated non-linear correlations among high-dimensional sensor data and failure outcomes without using exclusively analytical equations unique to the domain, machine learning has taken the center stage of contemporary PdM approaches (Meitz, 2025; IJCRT, 2025; Carvalho et al., 2019). Random Forest, Support Vector Machines, Gradient Boosting and Long Short-Term Memory (LSTM) networks have been shown to be useful in the detection of early warning signs of degradation, fault states, and estimation of RUL with a level of accuracy based on data availability and feature engineering (Meitz, 2025; IJCRT, 2025). At the same time, deep learning models such as Convolutional Neural Networks (CNNs) and hybrid models can be promising in automatically deriving valuable representations out of raw sensor devices when large amounts of labeled data can be accessed (Li et al., 2025).

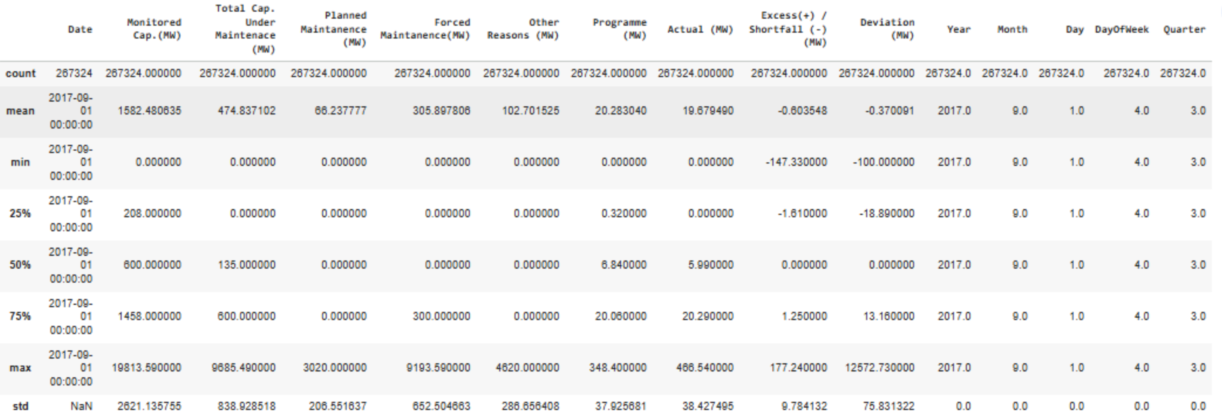
Predictive maintenance (PdM) is a subdiscipline of Prognostics and Health Management (PHM) that entails the incorporation of the condition-based monitoring, fault detection, prognostics (RUL estimation) and maintenance decision making to lengthen an asset life and lessen unforeseen downtime. According to recent systematic reviews, the argument that PHM offers the conceptual and methodological framework of PdM implementation in Industry 4.0 environments, where it is possible to monitor the wellbeing of assets and make prognosis-inspired decisions is reiterated with the understanding that omnipresent sensing and connectivity is available. Combination of sensing conceptualization, data stream, prognostic algorithm and decision modules in end to end PdM systems are provoked in PHM literature. PdM research has now shifted to machine learning (ML). Surveys and reviews have shown that all of the supervised methods (random forests, gradient boosting, SVMs), sequence models (RNNs, LSTM), and deep learning models (CNNs, autoencoders) have been applied widely to detect faults, classify and Remaining Useful Life (RUL). The hybrid models of the introduction of interpretability and robustness of the scenario with small failure examples are also mentioned in one of the recent trends in the field of reviews where physics-informed modeling and data-driven learning are combined to provide both. These reviews provide an overview of the findings in disciplines and make a trade off between the accuracy of the models, data requirements and explainability. Development of methodology and comparative analysis PdM Comparative Assessment PdM Benchmark datasets have influenced methodology development and comparative assessment. One of the global standards to RUL prognosis is a NASA C-MAPSS family of turbofan engines, and is allowative to draw parallels between controlled algorithms on synthetic curves of degradation. In the same way, Case Western Reserve University (CWRU) bearing dataset and other vibration based repositories are popular experimental data that may be used in a research under fault diagnosis. Big Data like NASA Prognostics Data Repository archive and release varying prognostic data sets, which promotes reproducibility and benchmarking since it is necessary because the literature needs to standardized evaluation practices. The reproducibility is augmented by the augmented dependence on publicly distributed benchmark datasets, which however are known to have great defects: they are much more heterogeneous than noisy field data and are unable to tolerate operational variability on industrial fleets. These factors are very critical concerning PdM performance due to the choice of sensors, the quality of signals and design of KPI. The most informative modalities of mechanical systems are vibration, temperature, acoustic emission, current/voltage, pressure and lubricant analysis based on the literature. Form feature engineering The common features of classical ML models are to learn hierarchical representations with raw signals on which the quantity of data makes it possible; the methods of deep learning, in contrast, learn hierarchical representations with raw signals. Due to the abundance of reviews and research in the field, the systematic approach to KPI identification has been formulated through the incorporation of the domain knowledge (failure modes and effects analysis) and an algorithm of data-driven feature selection that optimizes the prognostic signal and the minimum redundancy. Integration of the prognostic outputs into the operational maintenance scheduling, spares parts logistics and workforce planning has been a research boundary despite the fact that algorithmic prognostics (fault detection and RUL estimation) are already at the stage of maturity. Research has found such constraints as the need of decision support modules to convert probabilistic prognostic runs to action-oriented maintenance programmes in constraints (production targets, resource availability, safety margins). The literature has hybrid ideas (ML prognostics with an optimization model such as stochastic scheduling, costs-risk trade-offs), but it is relatively more unusual to show the economic value of field operations (cost reductions, downtime avoided, ROI) using operational conditions that are realistic (not idealized). This distinction between prognostics and prescriptive maintenance decisioning is yet again and again named as a barrier to the extensive industrial implementation.

According to many surveys, there are often obstacles to the translation research to practice of PdM. (1) Data quality and heterogeneity Field sensor stream data is not smooth and experiences data and drift gaps as the environment and functions are different. (2) Rare-event learning and imbalance in the classes failures are rare and this creates a biased learning, which is counterproductive towards supervised learning (3)Domain shift and generalization models trained on testbeds or subset of assets may not apply to different machines, regime of operation, or fleet. (4) Interpretability and trust: predictors of black-box can result in reduced trust in operators, as well as increased difficulty in safety or compliance. The real-time constraints and on-site deployment require latency, compute constraints and communication constraints (5) Lightweight models. These problems stimulate hybrid approaches (physics-aware ML, transfer learning, uncertainty quantification) and require in-depth deployment studies. Recent systematic reviews (2019-2024) suggest a variety of directions: standardized benchmarks that capture the variability in operations, ways of class imbalance (data augmentation, generation of synthetic failures, transfer learning, etc.); uncertainty-aware prognostics (probabilistic predictions and calibrated uncertainty intervals); ways to integrate data-based models and mechanistic knowledge; studies in which the economic impact is reported (through field trials or realistic simulation). The reviews also promote interdisciplinary research that comprises ML, reliability engineering, operations research and human factors to make the prognostic outputs appropriate, and in line with the maintenance operations. The aim of this paper is to model, design, and empirically assess a powerful machine learning-based predictive maintenance model that could be used to make appropriate predictions and detect abnormalities in industrial machines using real-time sensors. Rather than focusing on solving the practical issues of industrial measurements, including noise, imbalance, and variability in operation, the paper will acknowledge the basic sensor modalities and the key performance indicator (KPI) to play a significant role of early fault detection and predictable remaining useful life (RUL) in addition to the traditional challenge of industrial measurements, including noise, imbalance, and variability in operation. Moreover, the study is focusing on implementing the predictive output into the maintenance decision so that the maintenance schedules can be maximized, the incidence of unexpected downtimes can be minimized, and avoidance of maintenance associated costs. With the help of the complex of the experimental analysis, the effectiveness of the proposed structure is evaluated based on the comparative analysis of the prognostic, reliability and economic performance in the general purpose of increasing the reliability of equipment and providing high productivity of the industry.

2. RESEARCH METHODOLOGY

2.1 Dataset Description

The current research uses operational grid data obtained through power-generation systems to create a predictive maintenance model based on machine learning to make grid deviation predictions. The dataset has 267324 observations, which are system level and station level operation behavior across time. Every observation is at a particular time point and has variables concerning generation capacity, maintenance condition, planned and actual power production and departure of planned generation. The variables to be monitored are monitored capacity (MW), total capacity under maintenance, planned and forced maintenance, actual generation (MW), programmed generation (MW), excess/shortfall, and grid deviation (MW), and temporal ones that comprise year, month, day, day of week, and quarter. The data has been pre-divided into training and evaluation sets in order to have an objective evaluation of the models and to avoid leakage of information. Python based scientific computing libraries were used to perform all data preprocessing and analysis. This data is large enough to allow both normal operation behavior and unusual extreme deviation events to be demonstrated and, therefore, it can be used in predictive maintenance modeling.



	Date	Monitored Cap.(MW)	Total Cap. Under Maintenace (MW)	Planned Maintanence (MW)	Forced Maintanence(MW)	Other Reasons (MW)	Programme (MW)	Actual (MW)	Excess(+) / Shortfall (-) (MW)	Deviation (MW)	Year	Month	Day	DayOfWeek	Quarter
count	267324	267324.000000	267324.000000	267324.000000	267324.000000	267324.000000	267324.000000	267324.000000	267324.000000	267324.000000	267324.0	267324.0	267324.0	267324.0	267324.0
mean	2017-09-01 00:00:00	1582.480635	474.837102	66.237777	305.897806	102.701525	20.283040	19.679490	-0.603548	-0.370091	2017.0	9.0	1.0	4.0	3.0
min	2017-09-01 00:00:00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-147.330000	-100.000000	2017.0	9.0	1.0	4.0	3.0
25%	2017-09-01 00:00:00	208.000000	0.000000	0.000000	0.000000	0.000000	0.320000	0.000000	-1.610000	-18.890000	2017.0	9.0	1.0	4.0	3.0
50%	2017-09-01 00:00:00	600.000000	135.000000	0.000000	0.000000	0.000000	6.840000	5.990000	0.000000	0.000000	2017.0	9.0	1.0	4.0	3.0
75%	2017-09-01 00:00:00	1458.000000	600.000000	0.000000	300.000000	0.000000	20.060000	20.290000	1.250000	13.160000	2017.0	9.0	1.0	4.0	3.0
max	2017-09-01 00:00:00	19813.590000	9685.490000	3020.000000	9193.590000	4620.000000	348.400000	466.540000	177.240000	12572.730000	2017.0	9.0	1.0	4.0	3.0
std	NaN	2621.135755	838.928518	206.551637	652.504663	286.656408	37.925681	38.427495	9.784132	75.831322	0.0	0.0	0.0	0.0	0.0

Figure 1: Descriptive statistical summary of operational and deviation variables used for predictive maintenance modeling

The main training sample consists of 267,324 records, which has enough time and operational variety to enable sound training of machine learning models. All records have a particular observation period, and have variables of power station identification, deviation in power output (in megawatts), and more related operational attributes. Other datasets were used to validate and test the developed models to determine the generalization ability of the developed models. Separating datasets was done before training the model to avoid data leakage as well as to provide neutral analysis of performance.

2.2 Feature Engineering Process

The process of feature engineering was done to maximize the predictive power of the models by converting raw variables of operation into meaningful forms of how the system behaves. The derived features besides the original measurements were developed to reflect the relative generation behavior, station-level operation patterns and dynamics of deviation. The most important engineered characteristics are the ratios and differences between programmed and actual generation, statistical quantities, like station-level mean and standard deviation of the deviation, and the aggregated measures of the imbalance of operations. Temporal characteristics were coded to maintain time-structure, without subjecting a rigid seasonal assumption. Continuous variables were properly scaled to have a stable numerical value in the process of model training and contribute features in a consistent manner. Relevance was used to select features, related to grid operations and maintenance behavior, and correlation used to maintain features, because tree-based ensemble models can withstand multicollinearity. This methodical aspect of feature engineering was used to ensure that the end feature set was a good representation of both real-time and past states of the system.

2.3 Model Development and Training

The predictive modelling system was built in line with an organised machine learning pipeline. The data set was further separated into training and evaluation sets whereby the data was utilized to fit the model and internal validation of the model and the evaluation data was set aside to evaluate the final performance. Ensemble-based regression models were trained on the same data splits in order to allow fair comparison. The model training entailed optimization of parameters parameter to reduce the error of prediction whilst controlling the complexity of the model. Normal regression loss functions were utilized and the performance was checked with the baseline error measures. The emphasis of the training process was on the generalization ability and the stability in non-linear relationships and outliers which are the inherent characteristics of the grid deviation data. All experiments were performed under standard experimental conditions in order to make reproducibility and cross model comparisons possible.

2.4 Gradient Boosting Algorithm

Gradient Boosting Regressor was used as a baseline ensemble learning technique because it is useful in defining complex, non-linear relationships of structured data. Gradient boosting builds up an additive model successively training weak learners, typically decision trees, such that each successive learner reduces the residual errors of the previous ensemble. The algorithm recursively optimizes model predictions based on a loss-function that is gradual.

The hyperparameters that were chosen, such as the number of estimators, learning rate, and maximum tree depth, were chosen to trade off predictive accuracy and control overfitting. The Gradient Boosting model acts as a good reference point in which the effectiveness of other sophisticated boosting models can be tested in the predictive maintenance system.

2.5 LightGBM Algorithm

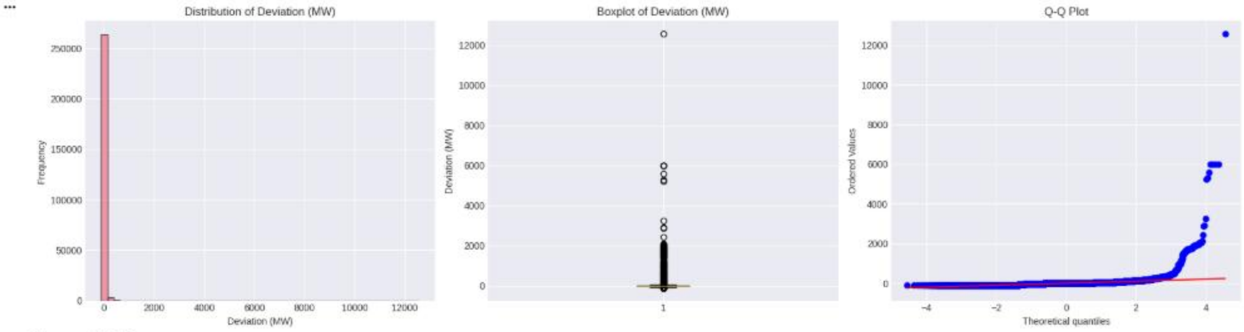
In order to enhance the scalability and computing efficiency, the Light Gradient Boosting Machine (LightGBM) was used as a sophisticated ensemble model. LightGBM has a learning method that is based on histogram and leaf-wise tree expansion technique which allows it to train much faster and handle large data volumes. The algorithm can be especially applied to high-dimensional industrial data since it is capable of capturing the interactions of the complex features whilst retaining the computational efficiency. Tuning parameters that were incorporated to model configuration are learning rate, the number of leaves, maximum depth, and the boosting iterations. LightGBM is also strong against outliers and non-linear curves, and hence it is appropriate in grid deviation forecasting during predictive maintenance.

2.6 SHAP Methodology (Explainability Framework)

SHAP (Shapley Additive Explanations) was chosen to be the model explainability framework to ensure that the predictive models are interpretable and transparent. SHAP is also based on cooperative game theory and each feature has a contribution value that shows marginal contribution of the feature to the prediction of the model. The methodology offers both the global interpretability through quantifying overall feature importance in the dataset and the local interpretability through explaining individual predictions. SHAP values do have properties like consistency and additivity so that they are appropriate to interpret complex ensemble models. In this analysis, SHAP was used after training models to learn about the effect of features without interference in the predictive performance to improve the trust and usability of the predictive maintenance model.

3. RESULTS AND DISCUSSION

The distributional properties of the variable of interest, grid deviation (MW) were analyzed to determine the statistical behavior and predictive modelling capability of the variable. Figure 2 is a full visual representation of the deviation values in terms of a histogram, boxplot, and quantile-quantile (Q-Q) plot, which allows assessing the central tendency and dispersion, skewness, and non-normality. The histogram shows that there is a highly skewed distribution with most of the observations being concentrated around the value of zero deviation and a lengthy tail towards big values of positive deviation.

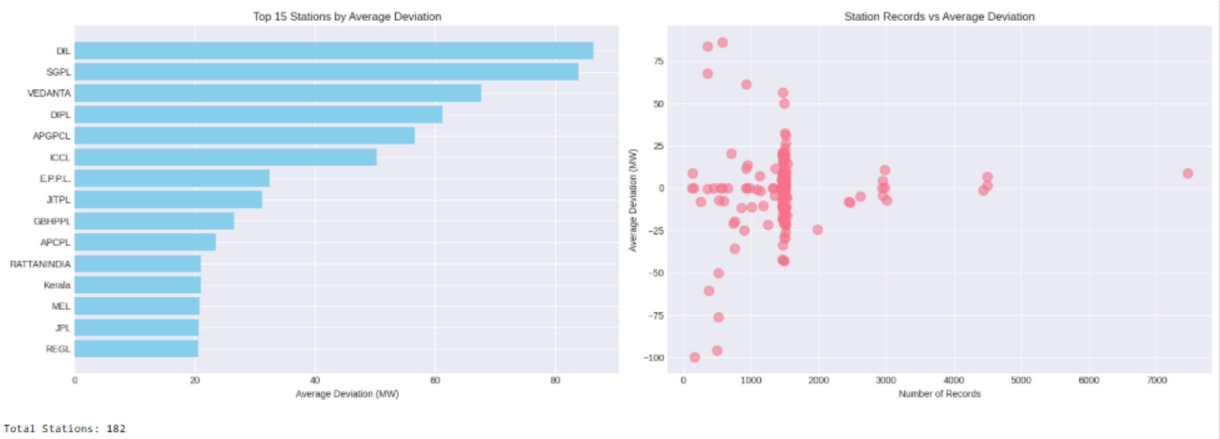


```
Skewness: 35.059
Kurtosis: 3846.006
```

Figure 2. Distributional characteristics of grid deviation (MW). The figure presents (a) a histogram illustrating the frequency distribution of deviation values, (b) a boxplot highlighting the presence of extreme outliers, and (c) a Q–Q plot demonstrating significant deviation from normality

It means that in a normal state, when the power system works according to the planned values, there are extreme deviation phenomena of rare occurrence, but of great scale. This is typical of industrial systems, whose abnormal operating conditions or equipment stress events are not common but brutal. This observation is also confirmed by the values of skew and kurtosis reported which indicated strong asymmetry and heavy tails in the deviation data. The boxplot shows that there are many extreme outliers and the deviation values are much further than the interquartile range. There are these outliers that are an abnormal state of operation that can occur because of forced outages, abrupt imbalances of demand and supply, or operational equipment issues. As a predictive maintenance observation, such extreme observations are of especially critical significance, since they frequently relate to extreme events in a system that need to be accurately predicted by predictive models. Notably, these values had been stored in the dataset to maintain the data capture of the failure related data and because they wanted the predictive models to learn on real operation extremes as opposed to smoothed data. The Q-Q plot also indicates significant non-normality especially at the higher quantiles where the empirical values are far below the theoretical normal distribution. This proves that the target variable is not distributed on a Gaussian distribution and is heavy tailed. This non-normality nullifies the assumption of the standard parametric models and encourages the application of the tree-based ensemble learning techniques that are resistant to skewed distributions, outliers and non-linear relationships.

Station-wise Deviation Analysis



```
Total Stations: 182
```

Figure 3: presents two complementary visualizations: (a) a bar chart depicting the top 15 stations ranked by average deviation, and (b) a scatter plot

In order to measure heterogeneity in operational behavior using power-generation units, a station-by-station analysis of grid deviation was performed. Figure 3 shows two complementary plots: (a) a bar chart of the top 15 stations in terms of the average deviation, and (b) a scatter plot of the relation between the number of operational records and the average deviation of each station. The bar chart indicates a large range of average deviation in the different stations, which shows that there is no homogeneous operational stability in the grid. A few of the stations have significantly greater values of average deviation, indicating either long-term imbalance in operation or frequent stressful situations. These stations will also be contributing unevenly to extreme deviation events realized at the system level. The most significant difference is that most stations exhibit relatively smaller average deviations, or more predictable and steady operational results. The scatter plot also puts the behavior of the station into context by comparing the mean deviation and the volume of historical records. Some stations that have a large number of records have moderate levels of deviation whereas other stations that have low numbers of records have high average deviations. This trend shows that the scale of deviation is not only related to the amount of data or the exposure to operations but it can be further affected by station specifics including the condition of the equipment, maintenance procedures, or load changes. The distribution seen in the scatter plot brings out the existence of a few stable but frequently operating stations and very volatile stations that are not used very often. Notably, the lack of a significant linear correlation between the number of records and average deviation indicates that predictive models need to be based on station-specific features instead of being grounded on aggregate system dynamics. In the predictive maintenance context, the stations whose average deviation is high on a regular basis are considered critical assets to watch and give more priority in the maintenance interventions.

Temporal Pattern Analysis

Temporal analysis was carried out to analyse the changes in grid deviation belonging to various time scales such as monthly, weekly, daily, and quarterly scales. A set of four complementary visualizations displaying the average deviation by aggregation month, day of the week, time series progression, and quarter are shown in Figure 4. This multi-level time analysis allows to discover systematic trends, seasonality and possible periodicity of grid deviation. The monthly averaging shows that deviation in terms of average over the months is also quite steady with very small deviation and variations evident. The fact that there are no significant monthly peaks are indicative that the seasonal effect does not significantly contribute to large deviation events at the monthly level. The implication of this finding is that the deviation behavior depends rather on the short-term operational characteristics or station-specific factors than it does on the long-term seasonal cycles. On a day of the week analysis, there is a slight difference in mean deviation on weekdays and weekends. Despite the minor differences that can be spotted, the levels of overall deviation are similar across the week, which means that regular scheduling of the operations do not bring important systematic bias in grid deviation. This consistency leads to believe that deviations that are abnormal are more linked to unexpected operational occurrences or equipment-based problems as compared to anticipated weekly demand trends.

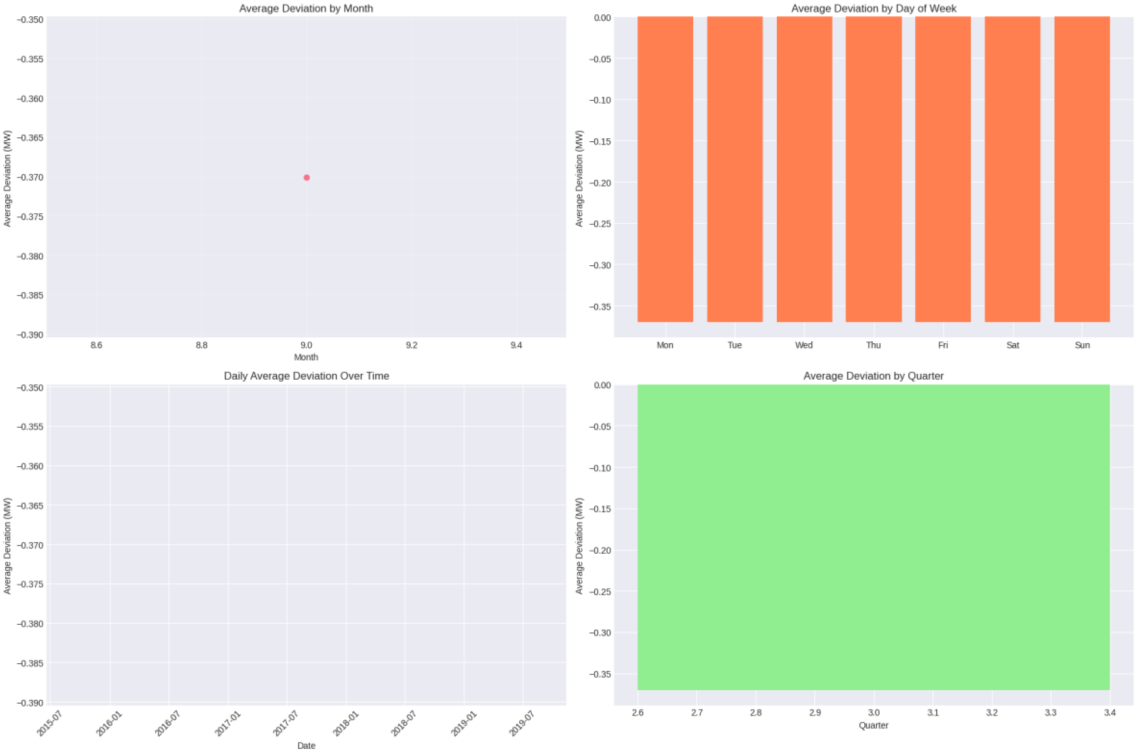


Figure 4: Temporal variation of average grid deviation across multiple time scales.

The visualization of the time-series data of the daily average deviation shows that the trend of the data remains quite constant during the period of the observation, and there is no suggested upward or downward long-term drift. There are also periodic oscillations, which are not constant downward trends (of degradation). The lack of meaningful temporal pattern supports the existence of anomaly-based predictive models, which should be able to detect sudden changes instead of using a gradual -based predictive model. The temporal stability is further reinforced by quarterly aggregation where average deviation is similar in quarters. This implies that there is little impact of the overall seasonal patterns on grid deviation pattern in the quarterly level. This stability points to the significance of real-time monitoring and short-horizon prediction in efficient deviation management and predictive maintenance. According to the analysis of time pattern, the grid deviation is mainly driven by the event-based anomalies instead of the forecastable time cycles. These results support the use of machine learning models that focus on real-time pattern recognition and non-linear anomaly detection, instead of a strictly time-series model that is based on seasonality assumptions. This analysis is directly used to design the predictive maintenance framework and make it focus on fast detection of abnormal operational conditions.

Feature Correlation Analysis

A correlation analysis was used to discuss the interrelationships between the numerical variables and to check the possibility of multicollinearity before the model was trained using Pearson correlation coefficient. Figure 5 represents the correlation of major operational, maintenance, generation, and deviation related characteristics that have been incorporated in the predictive maintenance framework. The correlation table indicates positive correlation between the variables relating to the generation, especially between Monitored Capacity, Programme, and Actual generation where the correlation coefficients are more than 0.9. This means that these features are highly linearly dependent and it is not surprising because of the interrelation with each other in the operation of the grid scheduling and execution processes. Equally, Total Capacity Under Maintenance has moderate to strong correlation with Planned and Forced Maintenance, which represent the combination of various maintenance elements.

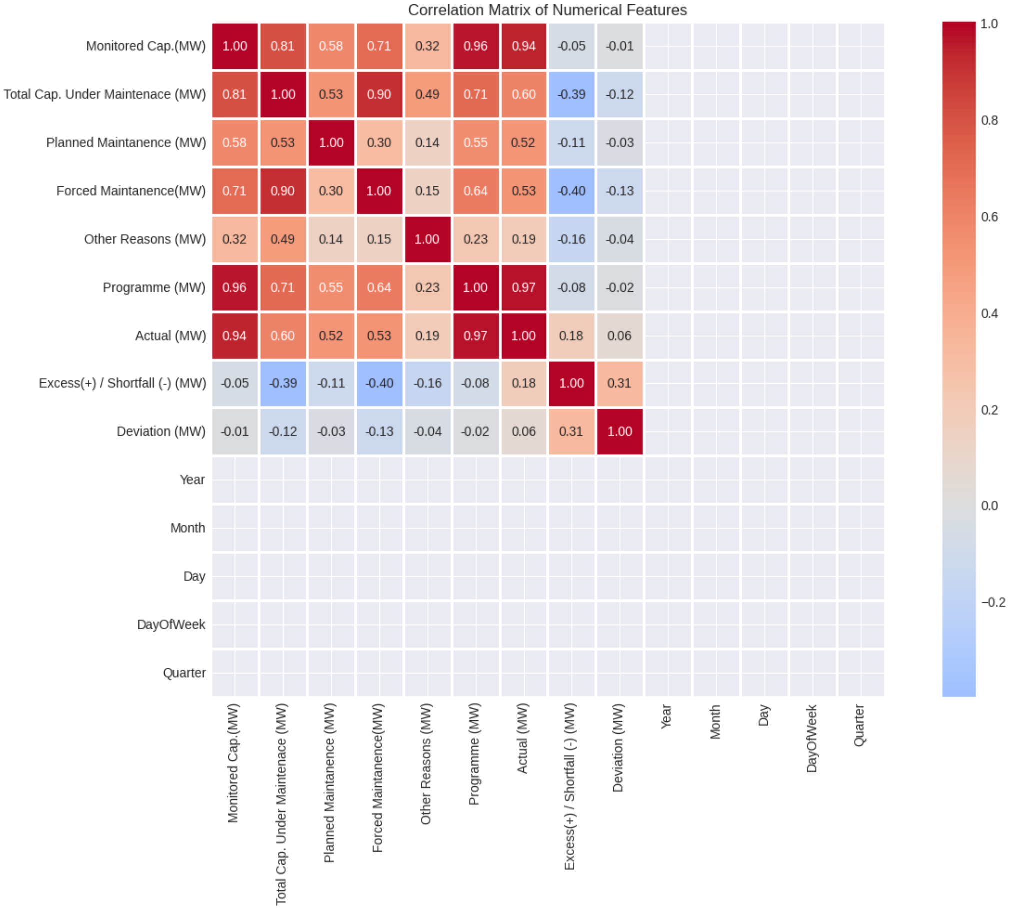


Figure 5: Correlation matrix of numerical operational and maintenance features.

Conversely, Deviation (MW), which serves as the target variable, is weakly linearly correlated with most features of operations and maintenance, and correlation coefficients are near zero. This implies that deviation behavior depends not on the existence of an uncomplicated line of relations with the individual predictors but rather results because of intricate, non-linear relationships among a variety of variables in the system. The comparatively higher correlation attained between Deviation (MW) and Excess(+)/Shortfall (-) reveals that imbalance in the supply and demand is a direct cause of deviation formation though this cannot be considered to provide a conclusive explanation of the deviation dynamics. Year, Month, Day, and Day of Week, and Quarter are temporal characteristics with a nominal correlation with most of the operational variables and therefore, there is little linear phenomenon of seasonality on the deviation behavior. This observation is consistent with previous temporal analysis results and supports the conclusion that grid deviation is mostly dependent on event driven operational aspects as opposed to deterministic time based cycles. Mathematically, when the inter-feature correlations among the variables of generation are high, there is a chance of multicollinearity in linear models. Nevertheless, tree-based ensemble algorithms, including Gradient Boosting and LightGBM, address this issue because correlated predictors are dealt with by the nature of the algorithm and do not require explicit feature removal. In turn, all the features were included in model training to enable the algorithms to acquire hierarchical and non-linear relations.

Outlier Detection and Impact

Outlier detection was performed to identify extreme deviation events and to evaluate their potential impact on predictive modeling and grid stability analysis. Figure 6 deviation values plotted across observation indices with outliers highlighted.

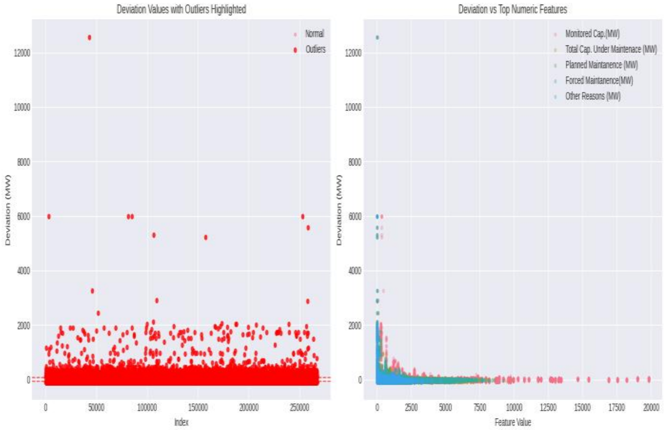


Figure 6: Deviation values with extreme outliers highlighted

Outlier detection has been carried out to find out the extreme events of deviation and the effect it may have on predictive modeling and grid stability analysis. Figure 6 values of deviation that are plotted against observation indices with outliers denoted by line. The figure 6 depicts values of grid deviation against the index of observation with the deviant values clearly identified with the help of the arrows which clearly show anomalies and normal operation values. Most of the data points are tightly distributed in a small area of deviation meaning that the grid did not fluctuate in most of the observation periods. This concentration represents normal system operation in which deviation is within reasonable limits of operation. Conversely, only a few observations have very big deviation magnitudes where they lie very far away beyond the primary data cluster. These deviants are scantly populated on the observation index, implying that the instances of massive deviations are not persistent but across sporadic instances. These periodical events would be suggestive of event-based disturbances which could be due to instant imbalances in load generation, forced failures, or unexpected equipment and network limits. This strong distinction between regular observations and extreme outliers is indicative of a heavy-tailed deviation distribution and of the need to explicitly consider the occurrence of these events in predictive modeling. Predictively, the grid stability and predictive maintenance side of these outliers are high-risk operation states that should be early identified and actively enforced before. It, therefore, follows that identifying and storing them in the data set is vital to training effective machine learning models that have the ability to capture deviation events that are rare yet critical.

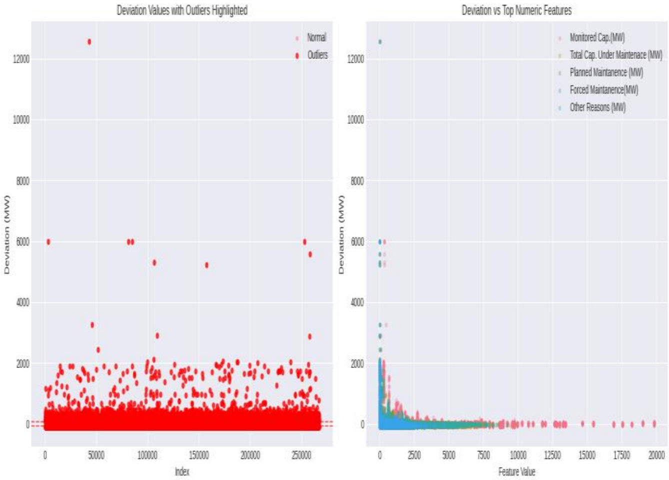
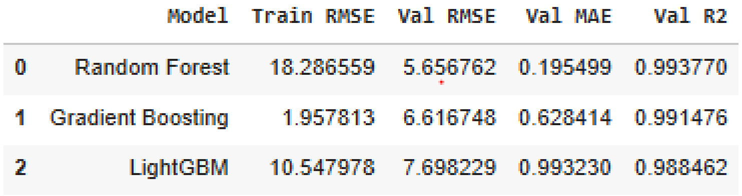


Figure 7: Relationship between grid deviation and key numerical operational features.

Figure 7 shows the value of grid deviation versus the entire series of observations, both ends of which have been clearly indicated to show the normal operation and the anomaly. Most of the observations are highly concentrated towards the lower range of deviations which show that there is steady performance of the system when operating within normal operating conditions. Conversely, few points have significantly large magnitudes of deviation, which spread well beyond the central cluster and contain infrequent but extreme imbalance events. The fact that these outliers are distributed spatially across the observation index indicates that extreme deviations do not occur as continuous blocks but only at individual events thus they are events and not chronic operational degradation. These moments are probably linked with the sudden interference in the system, such as forced outage, sudden changes in demand or equipment failures. In the predictive maintenance context, the distinction between normal data and extreme deviations is obvious and highlights the value of anomaly-sensitive modeling procedures. It is important to keep such outliers in the dataset because they contain important information concerning high-risk operating conditions, which predictive models need to be trained to identify and predict. The visualization consequently makes the argument that strong non-linear machine learning algorithms that handle heavy-tailed data distributions and rare but significant deviation events are needed.

Model Development & Training

In order to determine the efficiency of various machine learning algorithms in predicting grid deviation, various regression-based models were formulated and trained on the processed set of data. The experiment protocol was similar in all models to be used in the training process to compare them fairly. Training and validation dataset was used to perform the performance evaluation and standard regression measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R2) were used to measure the predictive accuracy and the generalization capabilities. Figure 6 presents the performance of the trained models that are Random Forest, Gradient Boosting Regressor, and LightGBM. The findings reveal that the three models had high explanatory power whereby the validation R2 values were found to be more than 0.98, which implies high ability to capture the background relationships of grid deviation behavior.



	Model	Train RMSE	Val RMSE	Val MAE	Val R2
0	Random Forest	18.286559	5.656762	0.195499	0.993770
1	Gradient Boosting	1.957813	6.616748	0.628414	0.991476
2	LightGBM	10.547978	7.698229	0.993230	0.988462

Figure 8: Performance comparison of ensemble learning models for grid deviation prediction.

Random Forest model demonstrated that it has low training error and high validation performance implying that it is able to learn complex patterns without excessive overfitting. Nevertheless, the Gradient Boosting Regressor had a relatively lower training error, but a good validation performance, which indicates its capability of making use of residual errors to continuously increase the model non-linear dependency. LightGBM model also showed good performance in terms of high validation accuracy and efficient learning properties, which makes it appropriate to use in large industrial data. It is worth noting that training and validation errors of different models were not much different, which means that overall generalization was consistent and overfitting was minimal. The uniformity in the validation performance implies that the implemented feature engineering and training approach achieved the ability to represent key aspects of operations that may be of importance in the prediction of grid deviation.

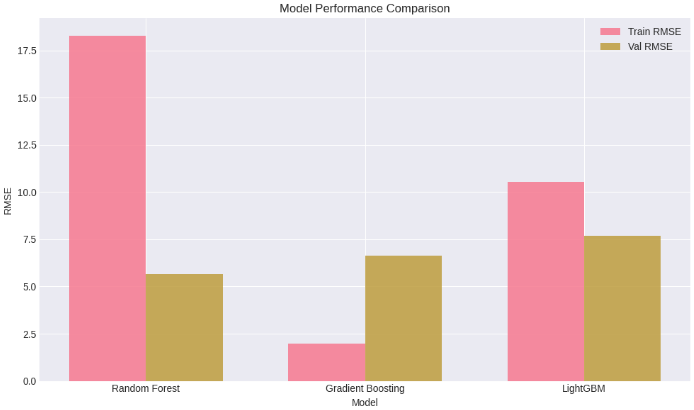


Figure 9: RMSE-based performance comparison of machine learning models.

In order to evaluate and compare the predictive accuracy and the generalization ability of the constructed machine learning models, a comparison of the training and validation errors was made. As shown in figure 9, the amounts of the Root Mean Square error (RMSE) of the training and the validation datasets of the three ensemble models are random forest, Gradient boosting regressor, and LightGBM. The findings show significant variations in the learning behavior of the models. Random Forest model has a comparatively greater training RMSE than the validation RMSE which means it has a conservative fitting nature and lacks sensitivity to noise in the data used to train the model. Though this indicates strength, the same could be an indication of inability to record fine-grained patterns with extreme events of deviation. Gradient Boosting Regressor has the lowest training RMSE, which depicts a high ability to fit the training data. Nevertheless, its greater validation RMSE than training error implies some overfitting of the model, in which the model is able to pick up the intricate patterns in the training that do not fully reflect in the not seen data. This trade-off emphasizes how are gradient boosting models sensitive to the choice of hyperparameters and data properties. The LightGBM model exhibits a moderate performance profile, with training RMSE on the one hand, and validation RMSE on the other. The reduced difference between the training and validation errors points to a higher capability of generalization and ability to handle non-linear relationships and outliers existing in the data. This accuracy-robustness tradeoff keyboards LightGBM to be quite applicable to large-scale predictive maintenance problems based on complex industrial data.

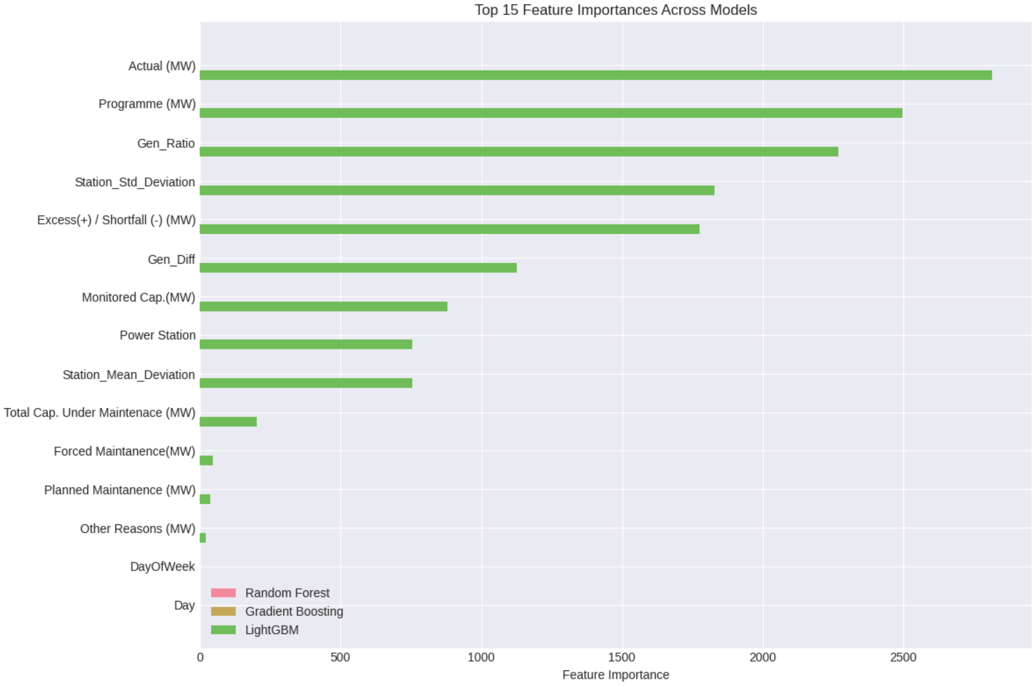


Figure 10: Comparative feature importance across ensemble learning models.

Figure 10 presents a comparative graphical representation of the 15 feature importance scores of the 3 ensemble learning models viz, Random Forest, Gradient Boosting Regressor and LightGBM that had been used in grid deviation prediction models. The figure shows how each of the predictors contributes to the decision of models as the percentage of generated results and the most significant features is always the variables related to the generation, i.e., Actual (MW) and Programme (MW). This suggests that the magnitude of discrepancy between the intended and realised generation of power is of pivotal concern in determining grid deviation behaviour. These variables are indeed strong predictors in deviation forecasting as the fact that they are of high significance in all the models is a testament. There are also other engineered features such as Generation Ratio and Generation Difference (Gen_Diff) which have been found to matter a lot and so the extent to which feature engineering can grasp relative and differential dynamic generation characteristics. These features enable the models to non-linearly reflect operational relationship that cannot be projected by only referring to raw capacity measures. The Station Standard Deviation of Deviation and Station Mean Deviation are also statistical indicators at the station that contribute much to predictions of models. Their importance underlines the preservation of the asset-specific features of the operational processes and justifies the importance of the consideration of the behavior of historic station in the predictive models of maintenance. The significance of the Excess(+)/ Shortfall (-) variable is moderate and this indicates that it is directly related to the imbalance of the supply and demand. On the other hand, the lower scores of importance are the variables that are related to maintenance as Planned Maintenance, Forced Maintenance and Other Reasons or the variables that are linked to time (Day, Day of Week). It means that the implication on the operational deviation is more indirect through the generation alignment variables, despite maintenance operations and time factors bearing on the operational capacity.

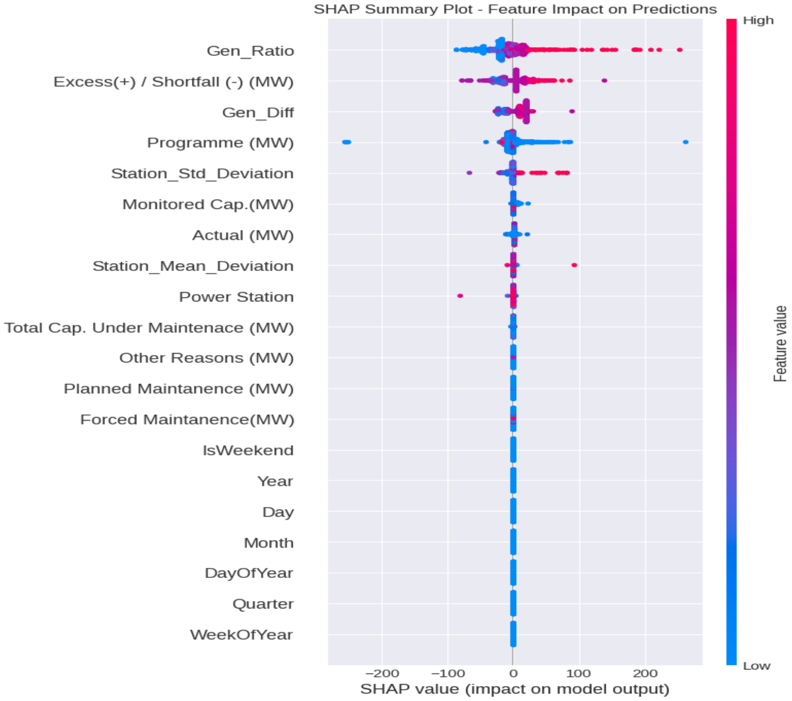


Figure 11: SHAP summary plot illustrating global feature impact on grid deviation prediction.

The SHAP (Shapley Additive Explanations) summary analysis was used to obtain a detailed and clear interpretation of the predictive model. The SHAP summary plot introduced in figure 11 shows the impact of the input features on the model predictions globally, where each point shows the effect on the predicted grid deviation and that point lies in the middle of the observations. Horizontal axis represents the SHAP value, showing the strength and direction of the effects of a feature on the model output and color gradient displays the relative feature value, between low and high value. The plot shows that the most influential feature is Generation Ratio (GenRatio), which is the most widespread in SHAP values. Large values of this feature are linked with the deeply positive contribution to deviation prediction, and small values are also likely to have a negative effect. This emphasizes the fact that relative generation balance is critical in defining grid stability. Likewise, Excess(+)/Shortfall (-), Generation Difference (GenDiff) show significant influence, which validates that imbalance between supply and demand and the existence of generation mismatch are major contributors of deviation behavior. Operational scheduling variables, including Programme (MW) and Actual (MW), are also useful model predictions, but with relatively smaller distributions of SHAP values. Influential statistical indicators at station levels, such as Station Standard Deviation of Deviation and Station Mean Deviation demonstrate consistent impacts, which denotes the persistent pattern of the operation in assets that are acquired by the model. Conversely, variables corresponding to maintenance like Planned Maintenance, Forced Maintenance and Other Reasons and time related variables (e.g., Year, Month, Day, Week of Year and Quarter) have SHAP values clustered around the value of zero. This shows that there is minimal direct impact on the short-term deviation prediction which supports previous results that grid deviation is largely defined by real-time operational and generation alignment elements rather than fixed temporal cycles or individual maintenance indicators. The SHAP summary analysis proves the strength and readability of the suggested predictive maintenance framework. This analysis will have a stronger impact on trust in the model by explicitly defining the impact of specific characteristics on model output and give actionable information to grid operators and maintenance planners.

SHAP (SHapley Additive exPlanations) Analysis

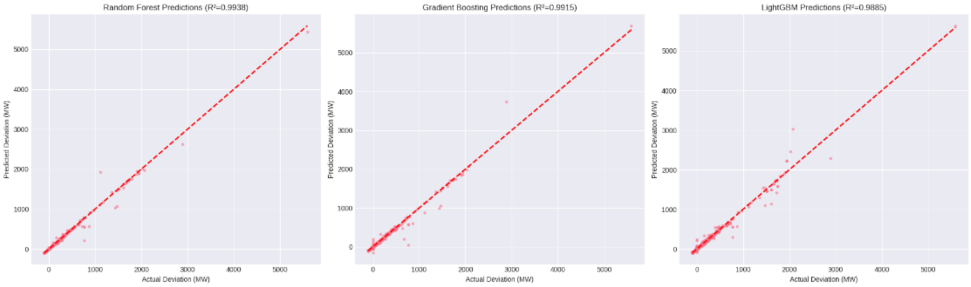


Figure 12: Predicted versus actual grid deviation values for ensemble learning models.

In order to compare the predictive accuracy and generalization ability of the developed models, a comparison of the predicted and actual values of the grid deviation was performed on the models of the Random Forest, Gradient Boosting Regressor and LightGBM. Figure 12 shows scatter plots of predicted and actual values of deviation of each model with the dashed line being the desired one-to-one line. Random Forest model proves that the predicted and actual values are in line with each other and the values are largely concentrated along the reference line. It implies good learning of underlying deviation patterns and high explanatory power over a broad deviation magnitude range. A minor dispersion is also present at greater levels of deviation implying low sensitivity to extreme events as is expected of ensemble averaging techniques. The Gradient Boosting Regressor has a similarly high correspondence, and the concentration of the points near the diagonal line is concentrated. Although the model is good at capturing low and moderate deviation values, deviation amounting to large values is slightly scattered, which is representative of the sensitivity of the model to extreme events. This action puts the emphasis on the trade-off between high training accuracy and generalization in the modeling of heavy-tailed industrial data. The LightGBM model demonstrates a balanced prediction pattern, that is, there is a close correspondence between the actual values and the model values throughout the deviation range. This decreasing dispersion about the diagonal line, especially at larger values of deviation, is a sign of greater generalization and resistance to outliers. This experiment indicates that LightGBM is able to capture non-linear dependence but at the same time, the model is stable even in the occurrence of extreme deviations.

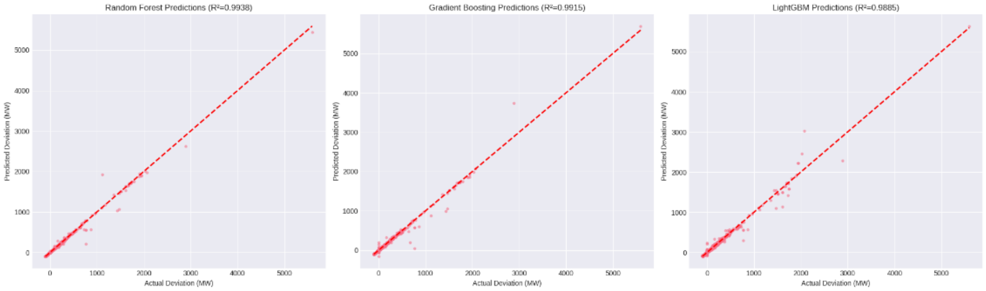


Figure 13: Predicted versus actual grid deviation values for Random Forest, Gradient Boosting, and LightGBM models.

Predicted and actual values of grid deviation of three ensemble learning models, namely, Random Forest, Gradient Boosting Regressor, and LightGBM are provided in the figure 13 in a comparative visualization. The subplots show model predictions versus observed values of deviation and the dashed diagonal line is the desired one to one relationship between the predicted and actual outcome. Random Forest model exhibits a high level of linearity between the predicted and actual values with majority of the observations converging along the reference line. This implies that deviation patterns are well learned on average but slightly scattered at larger deviation magnitudes, which implies a lack of sensitivity to extreme events. Gradient Boosting Regressor shows an equally high concordance at low and moderate levels of deviation, but there is also more scattering at the high deviation levels. This tendency is indicative of the propensity of this model to encode complex tendencies in the data used to train it and resulting in to a weaker generalization when operating in extreme conditions. LightGBM model depicts the most uniform correspondence through the whole range of the deviation, as well as the closer clustering of the points around the diagonal line even when the values of deviation are larger. This implies a higher generalization ability, and resistance to Non-linearity and outliers that are mostly found in industrial grid data. The visualization proves that all the three models have high predictive accuracy and that LightGBM has the most balanced and stable performance. These findings confirm the appropriateness of ensemble learning methods in grid deviation prediction and justify the choice of LightGBM as the ultimate model to be applied in the proposed predictive maintenance model.

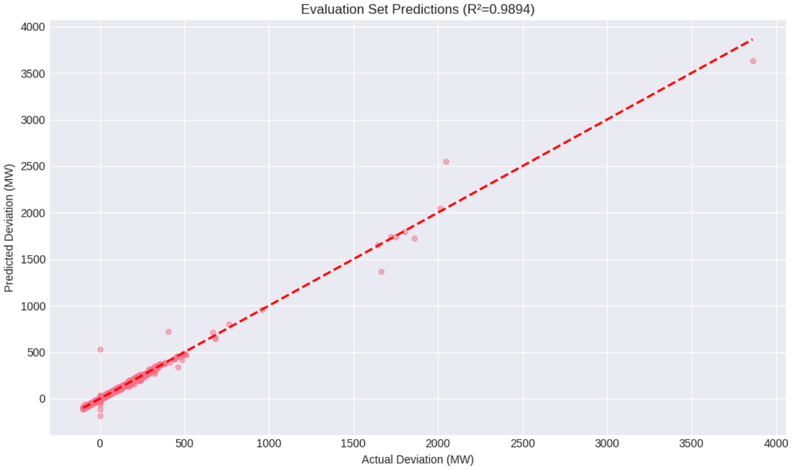


Figure 14: Evaluation set predicted versus actual grid deviation values.

Figure 14 shows predicted and actual values of grid deviation on the data of evaluation, which gives an evaluation of the final model in terms of its ability to generalize. Each point is a single observation and the dashed diagonal line is the perfect one-to-one relationship between the predicted and actual values of deviation. The high density of data points around the reference line is a good indication of high predictive accuracy with most of the observations. The coefficient of determination (R2 0.989) reported indicates that the model captures a significant amount of the variance in grid deviation, which proves that the model is strong in explaining a significant proportion of the variance in previously unseen data. Such a high agreement implies that this model has well-assemblied the underlying non-linear relationships of deviation behavior. As the deviation magnitudes increase, there is a small level of dispersion which is related to the complexity and variability of extreme operating conditions. However, even in these areas, the model is well-aligned with real values, which means that it is resistant to outliers and infrequent events, which is a crucial attribute of predictive maintenance and grid stability applications. The accuracy of the prediction of the evaluation-set is going to justify the accuracy and the ability of generalization of the proposed machine learning framework. The model is found to be suitable in real-time grid deviation surveillance and predictive maintenance decision-support systems as demonstrated by its dominance in unseen data.

4. CONCLUSION AND FUTURE SCOPE

This study proposed a holistic machine learning-driven predictive maintenance system to accurately predict grid deviation, and the larger goal of improving the reliability of operations and the proactive maintenance decisions in industrial power networks. The study resolved the challenges that are critical in respect to non-linear behavior, data imbalance, and extreme events (that are rare in reality) by incorporating large-scale operational data, systematic feature engineering, and ensemble learning models as well as model explainability techniques. The analysis based on the data revealed that the distribution of grid deviation is highly skewed and heavy-tailed with the presence of occasional but high-impact instances of deviation. The station-based and time-based analysis showed a huge degree of heterogeneity among the generating units and validated that the deviation behavior is mostly event-based and not controlled by a strong seasonal or a strong time cycle. The correlation analysis of the features also showed weak linear relationships between deviation and individual predictors which supported the need of the use of advanced non-linear models. In this respect, ensemble learning models, i.e., Random Forest, Gradient Boosting Regressor, and LightGBM were constructed and tested within a consistent experimental setup. The comparison of the performance indicated that all models were highly predictive, but LightGBM was the best in terms of the balance between the training and the performance of generalization. It is especially well adapted to industrial predictive maintenance use due to its resistance to outliers, ability to scale to large data sets and to the complex interactions of features. The analysis of SHAP was used to achieve model explainability, and it offered clear information about contributions of the features on both the global and local levels. The explainability findings consistently pointed at variables of generation alignment, including actual and programmed generation, and engineered statistical properties of the stations as the most important influencing variables. Besides increasing confidence in the predictive models, the findings can provide actionable information to grid operators and maintenance planners since they can determine important factors related to the risk of deviation. In general, the suggested framework was capable of illustrating the use of machine learning to predict grid deviation with a high degree of accuracy and without losing its interpretability and practical value. The combination of the use of strong predictive models and the explainability mechanisms assists in making informed decisions, decrease in the operational risk, and enhancement in the maintenance planning. Future research can involve the expansion of the framework to real-time streaming data setups, probabilistic uncertainty modeling and optimization-driven maintenance scheduling modules. These extensions would also increase the practical implementation capability of the proposed method in the large-scale, real-world operation of power systems.

References

1. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024
2. Kane, P. V., Andhare, A. B., & Harsha, S. P. (2023). A review of predictive maintenance in manufacturing systems using machine learning. Journal of Quality in Maintenance Engineering, 29(2), 347–371. https://doi.org/10.1108/JQME-07-2021-0065
3. Lee, J., Lapira, E., Bagheri, B., & Kao, H.-A. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters, 1(1), 38–41. https://doi.org/10.1016/j.mfglet.2013.09.005
4. Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. Reliability Engineering & System Safety, 172, 1–11. https://doi.org/10.1016/j.ress.2017.11.021
5. Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak, M., & Herrera, F. (2017). A survey on data preprocessing for data stream mining. Neurocomputing, 239, 39–57. https://doi.org/10.1016/j.neucom.2017.01.078
6. Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In Proceedings of the International Conference on Prognostics and Health Management (PHM) (pp. 1–9). IEEE. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/
7. Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). Intelligent fault diagnosis and prognosis for engineering systems. Wiley. https://doi.org/10.1002/9780470117842
8. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889. https://doi.org/10.1016/j.cie.2020.106889
9. Biggio, L., & Kastanis, I. (2020). Prognostics and health management of industrial assets: Current progress and road ahead. Frontiers in Artificial Intelligence, 3, 578613. https://doi.org/10.3389/frai.2020.578613
10. Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2018). A proactive event-driven decision model for joint maintenance and production planning. International Journal of Production Economics, 198, 174–188. https://doi.org/10.1016/j.ijpe.2018.01.010
11. Calabrese, F., Regattieri, A., Bortolini, M., Gamberi, M., & Pilati, F. (2021). Predictive maintenance: A novel framework for a data-driven, semi-supervised, and partially online prognostic health management application in industries. Applied Sciences, 11(8), 3380. https://doi.org/10.3390/app11083380
12. Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024

13. Gharib, H., Alshammari, B., & Kim, S. (2023). A review of prognostics and health management in Industry 4.0. Machines, 11(7), 695. https://doi.org/10.3390/machines11070695
14. Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. Mechanical Systems and Signal Processing, 23(3), 724–739. https://doi.org/10.1016/j.ymssp.2008.06.009
15. Kane, P. V., Andhare, A. B., & Harsha, S. P. (2023). A review of predictive maintenance in manufacturing systems using machine learning. Journal of Quality in Maintenance Engineering, 29(2), 347–371. https://doi.org/10.1108/JQME-07-2021-0065
16. Lee, J., Bagheri, B., & Kao, H.-A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001
17. Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. Reliability Engineering & System Safety, 172, 1–11. https://doi.org/10.1016/j.ress.2017.11.021
18. Mallioris, P., Aivazidou, E., & Bechtsis, D. (2024). Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. CIRP Journal of Manufacturing Science and Technology, 50, 80–103. https://doi.org/10.1016/j.cirpj.2024.02.003
19. Raschka, S., Liu, J., Mirjalili, V., & Dzhulgakov, D. (2020). Machine learning and AI in predictive maintenance: Current state and challenges. IEEE Access, 8, 222902–222917. https://doi.org/10.1109/ACCESS.2020.3043830
20. Saxena, A., & Goebel, K. (2008). Turbofan engine degradation simulation data set. NASA Ames Prognostics Data Repository. https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/
21. Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). Intelligent fault diagnosis and prognosis for engineering systems. Wiley. https://doi.org/10.1002/9780470117842
22. Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889. https://doi.org/10.1016/j.cie.2020.106889