MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3379-3386

# A Comprehensive Analysis of Machine Learning Models for Short-Term Wind Power Forecasting with Time-Series Validation and SHAP Explainability

Thummala Guru Sathvik, School of Computing, SRM Institute of Science & Technology,
Kattankulathur, Chengalpattu District - 603 203,
Tamil Nadu, India, Email id: gt0638@srmist.edu.in
MOHD Yaqoob Ansari, School of Computing, SRM Institute of Science & Technology,
Kattankulathur, Chengalpattu District - 603 203, Tamil Nadu, India, Email id: ma3652@srmist.edu.in
Dr. A. Pandian, Ph.D., PDF, Professor, Department of Computing Technologies, School of Computing, SRM Institute of Science & Technology,
Kattankulathur, Chengalpattu District - 603 203, Tamil Nadu, India, Email id: pandiana@srmist.edu.in
**Corresponding Author:** pandiana@srmist.edu.in

**Abstract**

Wind power forecasting is crucial for effective scheduling of energy grids and its integration with modern power systems. Power grid needs stable supply of power which enables in effective power management and helps in maintaining a balance between electricity supply and demand. Incorrect predictions may lead to power imbalance, grid instability and economic loses of the farm. However Wind power is highly variable as it depends on many atmospheric features which makes accurate WPF a challenging task. The data used for WPF is recorded in time stamps over the years with a particular time interval therefore it is important to consider the chronological order of the data while evaluating the models in order to avoid data leakage. However many existing studies performs random splitting, which results in data leakage and overestimated models. WPF has many non-linear relations like the time of the day which has it impact on all other features like temperature, wind speed and more but in a short-term period these atmospheric features will not change drastically making short-term forecasting more accurate. This study focusses on overcoming these challenges by implementing strict time-series split where different ML models where evaluated under both random and time-series split to find the importance of chronological order in future forecasting. Further more feature engineering techniques like lag features and rolling statistics were introduced which are used in one hour ahead WPF and SHAP analysis is used for interpretation of various features importance on the forecasting accuracy. Furthermore walk-forward validation is performed to evaluate the model robustness across different temporal folds. The results of this study shows that by introducing feature engineering techniques under strict time-series evaluation the forecasting accuracy improved where an $R^2$ of approximately 0.94 is achieved. This study shows the importance of proper evaluation methods for accurate future forecasting and introducing interpretable modelling in developing WPF systems.

***Keywords:*** *WPF (Wind Power Forecasting), Time Series Split, Feature Engineering, SHAP Explainability, Walk-Forward Validation, Short-Term Forecasting*

## 1. Introduction

With the increasing population and rapid industrialization the global energy demand has increased significantly. The dependence on energy generated from fossil fuels to meet this demand is neither cost effective nor environmental friendly, as they release harmful green house gases which has serious effects on both environment and climate. With these increasing concerns many countries are moving towards a sustainable and environmental friendly energy sources. Renewable energy has emerged as a key alternative, in this clean and sustainable energy transition wind power has become one of the fastest growing renewable energy sources which is actively contributing to modern electricity grid.

However, wind power is highly variable as it depends on many atmospheric conditions which makes accurate WPF a challenging task. But it is essential for more stable and efficient grid operations which includes mainly load balancing and energy dispatch. Inaccurate predictions may lead to imbalance to the stability of the grid therefore increasing operational costs. From the physical wind power equation we can find that wind is directly dependent on cube of wind speed and also air density. This shows that small variations in wind speed results in drastic variation in wind power generation similarly from air density equation we can find that air density is directly proportional to pressure and indirectly proportional to temperature so with increasing temperature the air density decreases and vice versa which also have effect on wind power generation. From this we can note that wind power is highly nonlinear and many future dependent and hence there is a need for accurate future wind power forecasting model.

Machine learning models are widely used in wind power forecasting than traditional models as they capture the complex non-linear relationships between environmental features and power output more efficiently which results in more realistic forecasting. However many existing studies use random splitting for model evaluation which leads to overestimated model performance and unrealistic forecasting results from possible data leakage. Wind power is highly time variant as other features keep varying with time, hence strict time-series forecasting is crucial for accurate future forecasting. And it is also important to explore the temporal feature engineering such as lag features which captures historical power generation patterns.

To address these limitations, in this study various ML models were evaluated on both random and time-series split which enables to understand the possible data leakage effects on model forecasting. Further, more temporal feature engineering techniques such as lag features and rolling statistics were introduced to capture the historical dependencies in power generation. As these features help's in enhancing the model's ability to capture the short-term patterns more effectively as the atmospheric conditions will not change drastically in a short period of time these lag features enables the model to forecast the next hour power production by using current weather input and past power generation which helps models to learn short term dynamics and improve forecast accuracy. SHAP-based explainability is used for better interpretation as this shows the feature importance and it's contribution to the final output power generated and walk-forward validation is done to test the model stability and robustness over different temporal folds. These results shows that strict time-series split and introducing temporal feature engineering has improved forecasting accuracy where $R^2$ score of approximately 0.94 is achieved and highlights the importance of proper evaluation methodologies and interpretation for more reliable WPF systems.

### Objectives of the Study

1. To evaluate baseline ML models and compare their accuracy under random split and strict time series based split.
2. To examine the effect of strict time series split on the model accuracy and interpretation of importance of time series based split for realistic forecasting.
3. To examine how the introduction of lag features through feature engineering impact the accuracy of the models under strict time series split.
4. To evaluate the performance of models for short term wind power forecasting.
5. To validate the model stability and robustness of forecasting using walk-forward validation.

### Scope and Significance

The study focuses on improving short term WPF under strict time series split with various machine learning and deep learning models by introducing feature engineering particularly lag feature for one hour ahead forecasting and SHAP analysis which has significance in helping the operators for better and efficient grid scheduling and proper supply-demand balancing and hence integrating the renewable energy into modern power which has significant effect in avoidance for fossil fuels and its adverse effects on environment.

**2. Literature Review:**

[1] Abdelsattar et al. (2025) Evaluated various ML and DL models on a dataset of 40,000 observations mainly on environmental factors with best $R^2$ of 0.723 which is used as base paper in our study. How ever it was revealed that they have done random split of 80:20 for evaluation of models but in true future forecasting strict time-series split is crucial as it avoids data leakage and the study have not explored any temporal feature engineering variables. The LSTM model used in this study was tested under random split which is not a right methodology for evaluation of LSTM which mainly depends on chronological order. Similarly [2] Karaman (2023) has evaluated various models on wind turbine SCADA dataset but has followed same 75:25 random split which raises the concern of data leakage and the study did not implemented lag features or validated the accuracy across various temporal folds.

[3] Rajaperumal and Columbus (2025) proposed a three stage forecasting framework by evaluating various models were evaluated and implemented hyperparameter tuning using grid search with five-fold cross validation and introduced a stacking modal of reported accuracy of $R^2$ of 0.998 which is really high in real time WPF when clearly examined this was found that they have used random split of 70:30 for model evaluation which raises the same concern of data leakage and no SHAP explainability. In the current study it addresses this by implementing strict time series split and feature engineering and walk-forward validation of models on different temporal folds.

[4] Dmitrijevs et al. (2025) conducted a comparative analysis on short-term wind energy forecasting using operational turbine data from Latvia and other datasets. A key finding of this study is that they performed modal evaluation in the horizon of 1 to 36 where $R^2$ of approximately 0.95 was achieved which gradually decreased to 0.14 by 36 hour. However they did not implemented lag feature engineering and not validated on different temporal folds. [5] Haq et al. (2025) they had performed a comprehensive review of various models in WPF from 2006 to 2025 which confirmed that ultra short term and short term forecasting has achieved lower error values due to the reduced in dependency on atmospheric conditions as mentioned earlier the atmospheric conditions will not change drastically in a short interval of time which is an advantage for accurate short-term forecasting. Furthermore they identified introducing lad features as a technique for improving the accuracy under time series predictions which was implemented in the current study.

[5] Haq et al. (2025) have done a comprehensive review of ML models where it was confirmed that tree based ensemble models have given strong generalization for short term forecasting. [6] Zaid Allal et al. (2025) have studied WPF at short horizon of 15 minute which achieved 50% improvement in RMSE that shows combining temporal features and autocorrelation informed features significantly improved forecast accuracy. However they did not include SHAP explainability analysis or walk-forward validation and mainly focuses on turbine operational parameters.

[7] Zhang et al. (2026) proposed a SHAP-guided Mixture of Experts framework for day-ahead wind power prediction using 16 meteorological features from a farm located in China. They have used SHAP values to guide the gating network of the ensemble model rather than just a interpretation tool which is a clear advancement in forecasting and shows the importance of SHAP in accurate forecasting. [8] Liao et al. (2024) conducted a study to evaluate four explainable AI techniques that include SHAP, Permutation Feature Importance, Partial Dependence Plots and LIME for WPF models and have confirmed that SHAP is the most trustable technique for global interpretation of tree-based models. They found WS100 as the primary influential feature in the dataset used by them. The current study demonstrates the SHAP analysis after performing the feature engineering by introducing lag features.

[9] Ahmet Durap et al. (2025) which applied explainable deep learning techniques for WPF in costal regions of Australia. This study shows that how SHAP analysis helps in finding the most influencing features in forecasting and model interpretability. Wind direction and humidity have found more influencing features for the particular dataset used in their study. However this study mainly focuses on wind speed prediction rather than wind power forecasting.

[10] Yang Yang et al. (2024) done a systematic survey of WPF techniques published in the Web of Science database over two decades, which confirms that ML based models have outperformed the tradinal models which were used before fore forecasting. ML models learns more complex patterns and the non-linear relationships between the features. LightGBM, XGBoost and LSTM were among the widely adopted models for forecasting. The current study uses these best performing tree based models for short term wind power forecasting with SHAP analysis and different temporal folds validation.

**2.1. Research Gap:**

Although existing literature provides valuable insights into wind power forecasting and evaluating strategies, several research gaps remain. First, many studies use random data splitting techniques for model evaluation, despite wind power strongly dependent on time. This approach can lead to overestimated models due to data leakage, where future information in used in the training set which enables the model to learn patterns and evaluated on test set which results in unrealistic forecasting performance. Second, when the forecasting is done truly based on environmental variables it is crucial to explore temporal feature engineering techniques such as lag features and rolling statistics which help the model to understand the past power values, particularly capturing short-term patterns to predict the future power generation under strict time-based forecasting. Third, many studies used simple train-test split for model evaluation and haven't done testing across various time periods which shows the model stability and robustness across different temporal folds. Techniques such as walk-forward validation are necessary in testing the model consistency. And, finally limited attention is given to model interpretability, while many studies have achieved high accuracy but often lack explainability making it difficult to understand each feature contribution on model accuracy. Overall, the literature suggests that integrating feature engineering techniques under strict time-series evaluation for short-term forecasting has significant impact which is because of the less variation in the atmospheric conditions in a short period of time making this an advantage for accurate forecasting. Therefore, further research is required to develop comprehensive framework that enable effective forecasting of future power generation by integrating strict time-series based evaluation, temporal feature engineering, robust validation techniques, and explainable models. Addressing these gaps is important particularly for short-term WPF, where accurate predictions are essential for grid stability and its efficient operations.

**3. Research Methodology:**

**3.1. Research Hypotheses:**

Based on the objectives of the study, the following null hypotheses were formulated:

$H_{01}$: Strict time series split based evaluation does not produce any different forecasting results compared to random split based.

$H_{02}$: Lag based temporal features have no significant effect on forecasting accuracy of models.

H03: Machine learning models has no significant effect compared to baseline models.

During the study different ML models were used for evaluation and they were modelled on both random split and strict time series based split where the change in accuracy was observed. Under time series split the accuracy dropped indicating random split causes data leakage causing increased accuracy of the model.

Lag based temporal features have significant effect on the model as this data is limited to atmospheric variables feature engineering helps the model to predict the future target value from current input and past power generated increasing the overall accuracy of the model.

Machine learning models has significant effect as wind power forecasting is highly volatile as it will not be constant and changes with atmospheric variables hence accurate and realistic power forecasting helps in grid stability. ML models can capture this complex non-linear relationship between different features than traditional models which helps in realistic accuracy improvement.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3379-3386

**3.2. Data Preprocessing:**

The dataset used in this study was first prepared well by performing a series of preprocessing steps which includes the conversion of time column of the dataset into datetime format which handles temporal information more effectively. Then the dataset in then checked for its correct chronological order to make sure the observations are in sequence which prevents data leakage during model training and evaluation, where the model learn from the past data and predicts the future values. And then the dataset is examined for all missing values and any inconsistencies in the time range mentioned in the dataset and large gaps in the data were dropped and additionally duplicate values were identified and removed. These preprocessing steps ensure clean, structured, and chronological ordered dataset, which is essential fore accurate WPF in this study.

**3.3. Feature Engineering:**

Feature engineering has a important role in improving the performance of wind power forecasting models. In this study various temporal features where introduced to capture the time dependent nature of the wind power. Lag features were created to store the power generated in the past hours which helps the model to understand the pattern in the power generation. Among these lag features the power lag of previous hour has a significant impact on the power produced in the next hour as in short-term forecasting the atmospheric conditions will not have a significant change. In addition to these lag features, rolling statistics were introduced which provides smoothed representation of recent power generated in previous hours by removing noise. As we know wind power is highly variant and with the changing atmospheric conditions which changes throughout the year hence time-based features like hour, day, and seasonal attributes were extracted from the datetime column of the dataset as this data span over years which helps in capturing variations in wind patterns over time. This features help the model in better capturing the complex non-linear relationships resulting in better accuracy in future forecasting under strict time-series evaluation conditions.

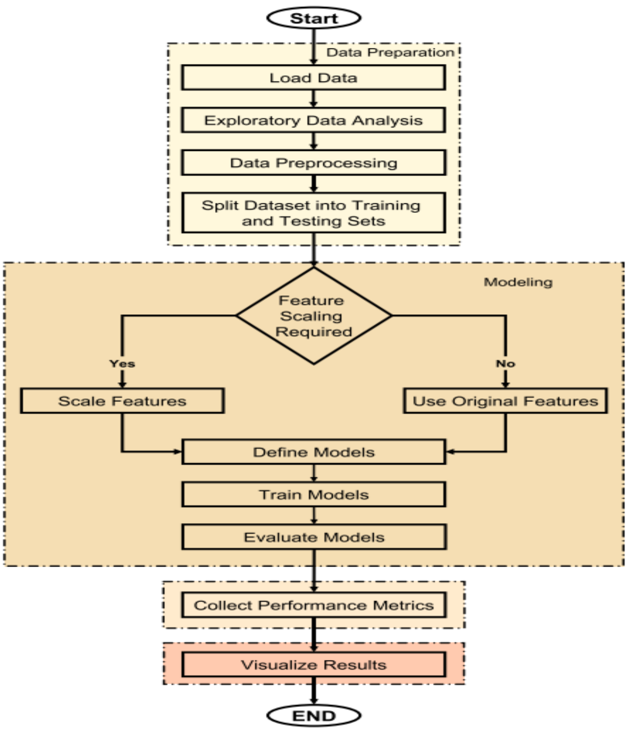
**3.4. Model Development:**

Various models were tested in this study under different training and testing split conditions of the data to compare their performance. The machine learning models used in this study include, Linear Regression, Support Vector Regression, Random Forest, ExtraTrees, XGBoost, LightGBM, CatBoost. These models were selected to capture both linear and non-linear relationships and patterns present in the data. Firstly, all these ML models were evaluated using random splitting of 80:20 for training and testing set and further the data is arranged in chronological order and a strict time-series split is done to evaluate these models. From the analysis it was found that random splitting is not suitable for future forecasting as it leads to data leakage by using future information in training set, resulting in unrealistic forecasting. Where as under strict time based split ensures the models are trained on past observations and then tested on future data thereby more realistic forecasting performance. Multiple models were used in this base line analysis because the Tree-based models such as Random Forest, Extra Trees, XGBoost, and LightGBM are more efficient in capturing complex non-linear relationships while the simpler models such as Linear Regression and SVR works well in capturing linear relationships and baseline reference, finally finding the best suitable models to evaluate on one hour ahead future forecasting after introducing feature engineering techniques.

**3.5. Model Evaluation Metrics:**

The performance of the various ML models were evaluated under two different dataset split for study purpose and the results were evaluated using metrics like $R^2$ (Coefficient of Determination) which indicates the amount of variance explained by the model. This is the primary metric used where higher $R^2$ value indicates the model is performing better. MAE (Mean Absolute Error), which measures the average prediction errors of the models lower the MAE indicates better the modal accuracy. RMSE (Root Mean Squared Error), which measures the average of square root of squared errors helps in penalizing large errors more heavily and lower the RMSE better the modal accuracy. Together these metrics help in comprehensive evaluation of the various models forecasting accuracy under different experimental conditions.

## 4. Architecture Diagram



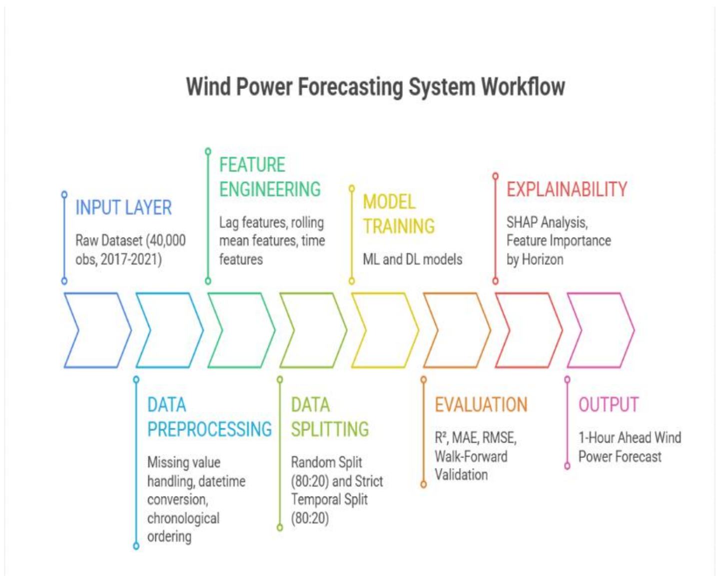




Fig. 1 – **Architecture Diagram**

Fig. 2 – **Data Workflow**

## 5. Data analysis & Interpretation:

### 5.1. Dataset Description:

This environmental dataset used in our study is part of a public dataset licensed under CC0: Public Domain, with the specific portion used in the study which has pure external atmospheric features which could be the external metrological data recorded in the location of the wind farm recorded over a period from 2017 to 2021. This contains wind speed at two different altitudes of 10m,100m usually wind turbine hub is placed in between 80m to 120m calculating the accurate wind speed at the exact height of the hub using internal sensors improves the accuracy of the forecasting. About 40,000 observations recorded with date and time stamp.

**Dataset Size:**

- Total Observations = approximately 40,000 and Time Range = ranging from 2017 to 2021
- Features include Temperature (2m), Relative Humidity, Dew Point, Wind Speed (10m), Wind Speed (100m),
- Wind Direction (10m), Wind Direction (100m), Wind Gust (10m),
- Target Variable: Wind turbine power output.

### 5.2. Descriptive Statistics:

The descriptive statistics of the environmental variables in the dataset are presented in Table 1 which provides a basic overview of the structure and variability of different features of the dataset used to evaluate various ML models. Key statistical measures such as mean and standard deviation were computed and it can be observed that wind speed at 100m has higher variability which has higher influence on wind power, further more the target wind power also has higher variability which shows the varying nature of the wind energy. The other atmospheric features do exhibit considerable variation which shows the dynamic behavior of environmental conditions. This statistical analysis enables us to observe the variability of wind power with variating environmental features and shows why wind power forecasting is a complex task and why there is need for a robust model for accurate WPF.

| Feature | Mean | Standard Deviation |
|---|---|---|
| **Temperature at 2m (degree celsius)** | **47.348** | **19.576** |
| **RH at 2m (%)** | **71.912** | **16.942** |
| **DWPT at 2m (degree celsius)** | **37.923** | **18.850** |
| **Wind Speed at 10m (m/s)** | **3.603** | **1.648** |
| **Wind Speed at 100m (m/s)** | **6.298** | **2.678** |
| **Wind Direction at 10m (degree)** | **203.639** | **96.985** |
| **Wind Direction at 100m (degree)** | **203.313** | **98.985** |
| **Wind Gusts at 10m (m/s)** | **7.811** | **3.576** |
| **Power Output (kW)** | **0.408** | **0.287** |

**Table 1. Descriptive statistics of environmental variables in the dataset**

### 5.3. Correlation Analysis:

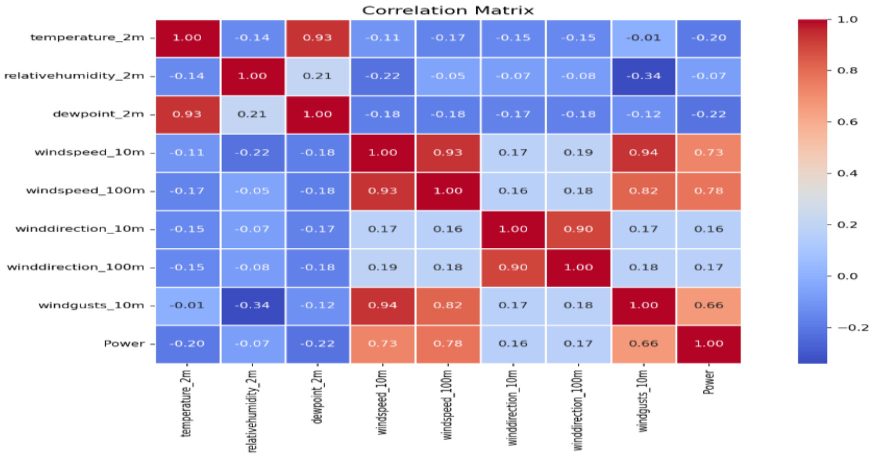



Fig. 3 – **Correlation Matrix**

The interactions between various features and their combined effect on the power output is analyzed using correlation matrix which is a color-coded heatmap. Power vs windspeed_100m: This has the strongest positive correlation with the resulting power output when compared with the other features. This shows that wind speed at 100m is much closer to the actual wind turbine hub hence it has high impact on the power generation.

Power vs windspeed_10m and Power vs windgusts_10m: These have moderate positive correlation with the power because the wind at 10m which is at surface level does not have much impact on the turbine and gusts contribute positively to power but are not as stable as steady wind.

Power vs temperature_2m, Power vs relativehumidity_2m, these have weak negative correlation with power showing that they are not reliable predictors.

Power vs dewpoint_2m, Power vs winddirection_10m, Power vs winddirection_100m have near to zero to weak negative correlation with wind power. This shows that they do not have meaningful linear relationship with power hence this shows they are not reliable predictors for with wind power forecasting.

### 5.4. Machine Learning Model Analysis:

In this study various ML models were used to predict wind power output using environmental variables, each of which is chosen for its specific advantage in regression task. The ML models includes Linear Regression, Support Vector Regression (SVR), Random Forest, Extra Trees, XGBoost, LightGBM, and CatBoost. The performance of these models where evaluated under random split and time-series split by using metrics like $R^2$, MAE, RMSE. Among these models the tree-based models like Random Forest, XGBoost, and LightGBM performed well with higher accuracy than other linear and traditional models as they are able to capture the complex non-linear relationships among features more effectively which results in higher accuracy. Further more significant difference is observed in the -

accuracy of models under random split and strict time-series split. Where the accuracy has dropped under time-series split which indicates that random split causes data leakage and overestimation of the model performance hence, this shows the importance of time-series split during accurate future forecasting of wind power. The results of different models under both random and time-series split are presented in the below figure Fig. 4 for understanding and $R^2$ comparison bar graph of various models is analyzed in Fig. 5.

| | Model | Split | R2 | MAE | RMSE |
|---|---|---|---|---|---|
| 1 | LinearRegression | Random | 0.621 | 0.1368 | 0.1743 |
| 2 | LinearRegression | TimeSeries | 0.6342 | 0.1419 | 0.1769 |
| 3 | SVR | Random | 0.6931 | 0.1184 | 0.1568 |
| 4 | SVR | TimeSeries | 0.6961 | 0.1244 | 0.1613 |
| 5 | AdaBoost | Random | 0.6261 | 0.1402 | 0.1731 |
| 6 | AdaBoost | TimeSeries | 0.5468 | 0.1636 | 0.1969 |
| 7 | RandomForest | Random | 0.7177 | 0.113 | 0.1504 |
| 8 | RandomForest | TimeSeries | 0.6592 | 0.1328 | 0.1708 |
| 9 | ExtraTrees | Random | 0.7262 | 0.1108 | 0.1481 |
| 10 | ExtraTrees | TimeSeries | 0.6662 | 0.1316 | 0.169 |
| 11 | XGBoost | Random | 0.7051 | 0.1167 | 0.1538 |
| 12 | XGBoost | TimeSeries | 0.6673 | 0.1307 | 0.1687 |
| 13 | LightGBM | Random | 0.7032 | 0.1178 | 0.1542 |
| 14 | LightGBM | TimeSeries | 0.6847 | 0.1278 | 0.1643 |
| 15 | CatBoost | Random | 0.7118 | 0.1153 | 0.152 |
| 16 | CatBoost | TimeSeries | 0.6813 | 0.1283 | 0.1652 |

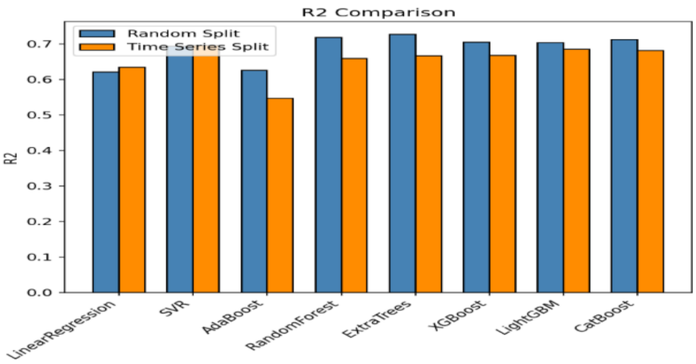



Fig. 4 – **Results of Machine Learning Models**

Fig. 5 – **R² comparison of Machine Learning Models**

Further, more temporal features engineering is explored which plays a crucial role in improving the performance of machine learning models for time-series forecasting. In this study temporal features like lag features and rolling statistics were introduced to capture historical patterns more effectively. Lag features represent the past power values such as Power_lag_1 which represents the power at the previous hour and rolling statistics, such as moving averages which provide a smoothed representation of recent values by reducing noise. These features are effective particularly in short-term forecasting, where atmospheric conditions do not change significantly. As a result these past power values serve as strong predictors of future power output. Further the best performing model like ExtraTrees which uses randomized decision trees which reduces the variance, LightGBM is a gradient boosting framework used for higher efficiency which performs leaf-wise tree growth and faster training on larger datasets, XGBoost which another powerful gradient boosting algorithm where used to forecast one hour ahead WPF.

| | Model | Horizon | Split | R2 | MAE | RMSE |
|---|---|---|---|---|---|---|
| 1 | ExtraTrees | 1 | TimeSeries | 0.9442 | 0.0364 | 0.0691 |
| 2 | LightGBM | 1 | TimeSeries | 0.9473 | 0.0378 | 0.0671 |
| 3 | XGBoost | 1 | TimeSeries | 0.9467 | 0.0367 | 0.0675 |

Fig. 6 – **R² comparison of Models over 1 hour ahead forecasting**

**5.5. Feature Engineering Impact Analysis:**

The auto correlation of wind power out is done which shows how strongly a variable or a target is related to its past values, in this case the target power to the lag features which represent previous hour power generation output. The lag features where represented along the x-axis in hours format and the y-axis shows the correlation strength. From this we can learn that Power_lag_1 is highly correlated with the target value this shows why 1 hour ahead forecasting is efficient we can see from the plot wher for first few hours the correlation is very high which gradually decreases this shows that in short term the variation in atmospheric conditions will not change drastically and hence the power generated in the past hour can be used to learn patterns to forecast the power generating in next hour. This shows hoe feature engineering effects the accuracy of the model even under time series split.

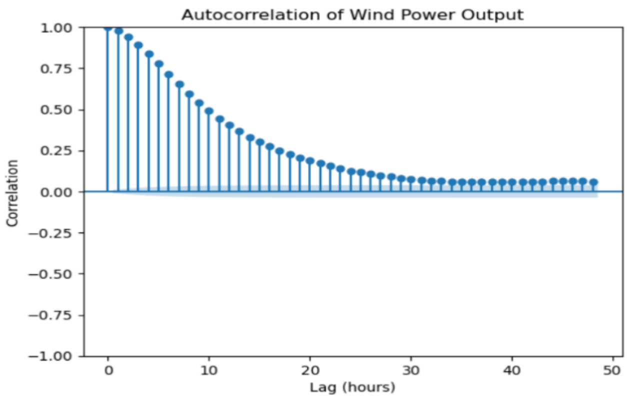



Fig. 7 – **Autocorrelation of Wind Power Output**

**5.6. Walk-Forward Validation:**

Further walk-forward validation is done to check the robustness of the models under various folds. The below plot shows the performance of the forecasting model across different temporal folds. The x-axis represents the validation fold number and the y-axis represents the corresponding $R^2$ score. This shows us the model stability under changing data used for evaluation. In this validation approach first the model is trained on a specific time period and then tested on the subsequent time interval

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3376-3375

and this is repeated by expanding the training window. This helps to find any variations in the model accuracy rather than a single train-test split. The use of walk-forward validation provides more realistic evaluation of the model consistency over different temporal folds and the results shows the model stability across different folds which shows its robustness.

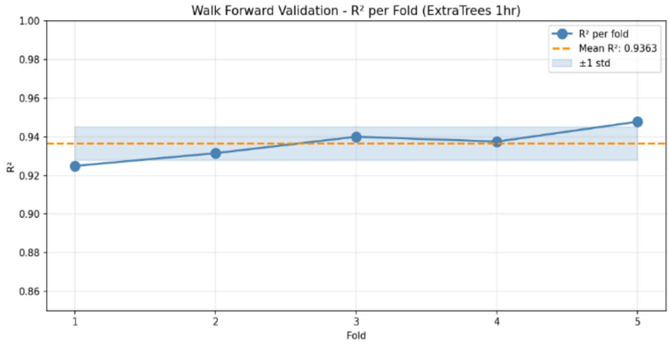



Fig. 8 **– Walk-Forward validation per fold**

**5.7. Explainability Analysis (SHAP):**

To improve the interpretability of the forecasting of machine learning models, SHAP (Shapley Additive Explanations) was performed which helps in evaluating the contribution of each feature towards the model forecasting accuracy. This is useful as we can analyze how each feature has its effect in the models output prediction. This shows the importance of various features that need to be considered during model evaluation. In the current study the SHAP summary plot reveals that the temporal features, particularly lag features have higher impact on the wind power output. In short-term wind power forecasting the next hour power generation strongly depends on the previous hour power generation as in a short interval of time there will be no significant changes in atmospheric conditions. The bar plot of mean SHAP values further confirms that the lag features dominate the model predictions which is followed by the key environmental variable wind speed at 100m which is ideally at the height of the hub of wind turbine.

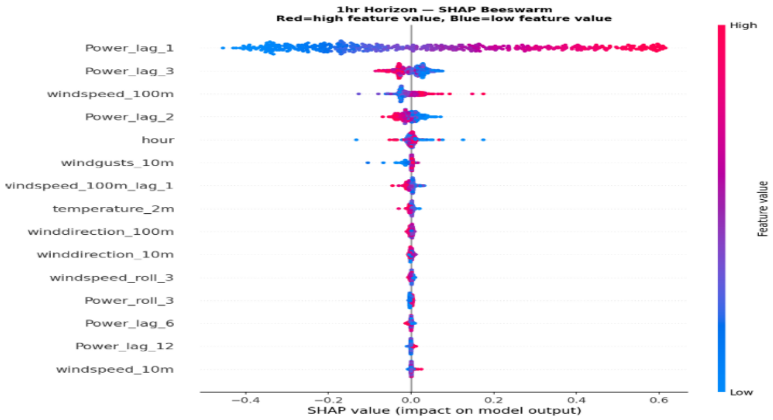



Fig. 9 **– SHAP analysis impact of features on output**

Further more the SHAP analysis is done for different forecasting horizons to find the feature contribution and dominance in model's output which is found that for one hour ahead forecasting lag features has the majority in contribution to the power output, however with the increasing time horizons the environmental features like wind speed has becomes more significant factor in determining model forecasting accuracy. This analysis shows that the temporal feature engineering helps in accurate forecasting in short-term WPF due to the stability of atmospheric conditions and hence this analysis helps in better interpretation of model performance.

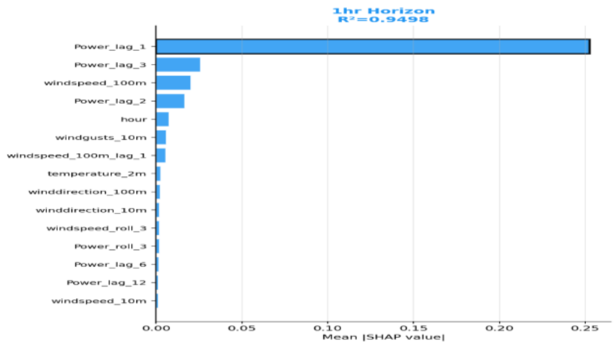



Fig. 10 **– Mean (SHAP Value) over 1hr Horizon**

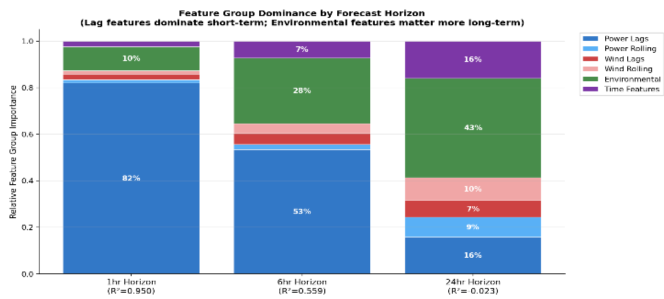



Fig. 11 **– Feature Group Dominance by Forecast Horizon**

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3376-3375

## 6. Discussion and Findings:

The results of this study shows the importance of machine learning models in wind power forecasting than traditional models as they are more efficient in capturing the complex non-linear relations between features. The Tree-based models have performed well which shows their ability in capturing the complex non-linear patterns effectively thereby increasing the model accuracy in forecasting. Another significant finding was that the data splitting methodology has a huge impact on model performance as various machine learning models were evaluated under both random and time based splits which shows the importance of time based evaluation in realistic wind power forecasting which avoids the problem of data leakage. This shows that it is important to follow chronological order during future wind power forecasting. Feature engineering was found to play a key role in short-term wind power forecasting which is further confirmed during SHAP analysis which shows how each feature contributes towards the final output and feature importance in accurate forecasting. Finally the model stability and robustness across various time folds was evaluated using walk-forward validation. Overall the findings of the study shows how proper evaluation methodologies, effective feature engineering and model interpretability will enhance WPF systems accuracy in forecasting which positively contribute to better grid stability and its operations.

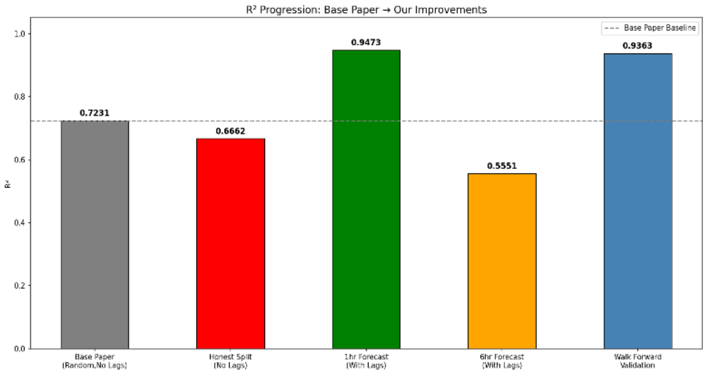



Fig. 12 – **$R^2$ comparison of Best Models under various scenarios**

The variation of model performance under different evaluation strategies and techniques is presented in the Fig. 12. The baseline model under random split has achieved $R^2$ around 0.72 however, under time-series this decreased which indicates data leakage in previous methods. With the introduction of feature engineering techniques particularly the lag features have improved the $R^2$ significantly to approximately 0.94 for one hour ahead forecasting and as the forecasting horizon increasing the accuracy decreases due to the lack of features and uncertainty. The walk-forward results further confirms that the model maintains stable performance over different temporal folds which validates the model robustness and reliability.

## 7. Conclusion:

The study focuses on realistic forecasting under short term interval where the variations in atmospheric conditions are limited and they do not change drastically in an give short interval of time the atmospheric conditions have great impact on the wind power generated as from the wind power formula we can note that the power generated is directly proportional to the cube of the wind speed hence small variations in wind speed will drastically effect the wind power generation making WPF a complex task however for stable balancing of the grid and to maintain a stability between the supply and demand scheduling of the grid is very important. Hence accurate forecasting helps to overcome these challenges up to an extant making energy supply stable and helps in maintain power grid stability by which renewable energies like wind energy can be integrated to the modern power systems.

The tree based models performed well on this large structured time series data for which accuracy of $R^2$ values close to 0.94 were achieved for one hour ahead forecasting and SHAP analysis is performed over time horizon to find the feature impact on model forecasting accuracy and which is further validated with walk-forward validation to evaluated the stability of the model and over all average accuracy of the model over different folds of the data under strict time series split. This shows us, tasks like WPF which includes chronological data a time series split is required for split of training and testing data on which models can be evaluated for realistic forecasting. Future research may explore advanced hybrid models, deep learning models such as LSTM and GRU for capturing long-term dependencies. Further improvements can be achieved by using larger datasets spanning over many years, and probabilistic forecasting techniques to further improve the model accuracy in wind power forecasting.

**References**

[1] Abdelsattar, M., Ismeil, M. A., Menoufi, K., AbdelMoety, A., & Emad-Eldeen, A. (2025). Evaluating Machine Learning and Deep Learning models for predicting Wind Turbine power output from environmental factors. *PLOS ONE*, *20*(1), e0317619.

https://doi.org/10.1371/journal.pone.0317619

[2] Karaman, Ö. A. (2023). Prediction of wind power with machine learning models. Applied Sciences, 13(20), 11455.

https://doi.org/10.3390/app132011455

[3] Rajaperumal, T. A., & Columbus, C. C. (2025). Enhanced wind power forecasting using machine learning, deep learning models and ensemble integration. Scientific Reports, 15, 20572.

https://doi.org/10.1038/s41598-025-05250-3

[4] Dmitrijevs, N., Komasilovs, V., Orlova, S., & Kamolins, E. (2025). Short-Term Wind Energy Yield Forecasting: A Comparative Analysis Using Multiple Data Sources. Energies, 18(16), 4393.

https://doi.org/10.3390/en18164393

[5] Haq, I. U., Kumar, A., & Rathore, P. S. (2025). Machine learning approaches for wind power forecasting: a comprehensive review. Discover Applied Sciences, 7, 1139.

https://doi.org/10.1007/s42452-025-07675-x

T

[6] Zaid Allal, Hassan N, Ola Salman, Khaled Chahine, Machine learning-based prediction model of wind turbine power generation, Cleaner Energy Systems, Volume 12, December 2025, 100218.

https://doi.org/10.1016/j.cles.2025.100218

[7] Zhang, H., Qin, G., Chen, X., Lu, L., Zhang, Z., & Song, J. (2026).A Day-Ahead Wind Power Dynamic Explainable Prediction Method Based on SHAP Analysis and Mixture of Experts. Energies, 19(1), 124. https://doi.org/10.3390/en19010124

[8] Liao, W., Fang, J., Ye, L., Bak-Jensen, B., Yang, Z., & Porte-Agel, F. (2024). Can we trust explainable artificial intelligence in wind power forecasting? Applied Energy, 376, 124273. https://doi.org/10.1016/j.apenergy.2024.124273

[9] Ahmet Durap. (2025). Explainable deep learning techniques for wind speed forecasting in coastal areas: Integrating model configuration, regularization, early stopping, and SHAP analysis. Neural Computing and Applications July 2025.

https://doi.org/10.1007/s00521-025-11433-w

[10] Yang Y,Hao L, Jinran W, Shaotong Z,Shangce G, A survey on wind power forecasting with machine

Learning approaches Published: 18 May 2024 Volume 36, pages 12753–12773, (2024).

https://doi.org/10.1007/s00521-024-09923-4