Grenze International Journal of Engineering and Technology, January Issue

Hybrid Statistical, Machine Learning, and Deep Learning Framework for Accurate Wind Power Forecasting using Big Data

Khushboo Pawar[1] and Dr. Devdas Saraswat[2]
[1-2]LNCT University, Bhopal, MP, India
Email: drkhushboopawar@gmail.com, devdass@lnct.ac.in

***Abstract*— To accurately predict wind power, the suggested study employs a comprehensive hybrid system that combines statistics, ML, and deep learning models. Improved prediction accuracy and model robustness are primary goals of effectively modeling large-scale, time-series data for both linear and non-linear connections. The most applicable meteorological variables in the 568,139 records contained in ten CSV files that constitute the Kaggle Wind Power Repository database are wind speed, direction, temperature, humidity, and air pressure.. Data preprocessing (cleaning, standardization, and extraction of temporal features) and dynamical model creation based on statistical models (ARIMA, SARIMA, SARImax), learning regressors (Random forest, XGBoost, LightGBM, SVR, El), and other methods make up the mentioned technique.**

The model's performance was validated and thoroughly tested using a number of metrics, including RMSE, MAE, MAPE, R 2, and NSE, which were evaluated using the experimental data. In comparison to more conventional methods of machine learning and traditional approaches, experiments have shown that deep learning models, and CNN-LSTM hybrid, in particular, do much better. CNN-LSTM was the best model to use when it comes to utilizing time trends and generate accurate and reliable predictions due to the best results in relation to the RMSE, MAPE, R2, and NSE. The results highlight the opportunities of hybrid deep learning models in enhancing predictive renewable energy and grid stability.

***Keywords*— Wind Power Forecasting, Machine Learning, Deep Learning, CNN-LSTM, Renewable Energy Prediction.**

I. INTRODUCTION

Pattern sequencing similarity is a potential solution to the prediction of wind power electricity generation since it capitalizes on the observed patterns in time series data to enhance better prediction. Particularly relevant to big data is the ability to sift through massive amounts of multimodal time series data in search of patterns that can inform future forecasts.

Pattern recognition and machine learning schemes are integrated to increase the predictive capacity of wind power production by absorbing the complicated two-way connections among various variables that affect the wind energy production.

Grenze ID: 01.GIJET.12.1.116
© Grenze Scientific Society, 2026

Some of the most important problems with accurately predicting wind power generation have an impact on the dependability and efficiency of renewable power systems.

- Data Variability and Complexity: Wind energy considerations are affected by the large number of meteorological factors and thus are difficult to predict because it is non-linear and non-stationary [1] [2]–[4].
- Data Acquisition and Quality: It is hard to get good quality high-resolution data-sets particularly in the complex terrains where sensor precision may be affected [5].
- Temporal and Spatial Dynamics: A non-periodic and erratic cycle of wind necessitates models that can be used to effectively capture the patterns of time and space(”Comparison and Enhancement of Machine [6].

Most traditional statistical models can't account for these kinds of differences, and cutting-edge machine learning techniques need massive amounts of high-quality data, which isn't always available. Moreover, there is also a spatial and temporal variability in various geographic areas which contributes another level of complexity because the local terrain, weather conditions and seasonal changes have a major influence on forecasting[7]. These issues do not only hamper the best integration of the grid but also raise the cost of balancing between the supply and demand. Therefore, the inability to overcome the constraints of wind power forecasting is an acute issue that should be mitigated to achieve better energy security, market competitiveness, and sustainable introduction of wind energy technologies. Machine learning has the potential to greatly improve predicting accuracy, but it is important to consider the limitations of these models as well, such as the requirement for massive datasets and powerful computers. Also, the use of new technologies such as IoT and smart sensors might also upgrade prediction systems in multiple ways, such as the provision of real-time information and the strength of the models.[8]–[11].

II. Literature Review

(Vaitheeswaran & Ventrapragada, 2019). Based on a training of an LSTM, or long short-term memory, neural network, this study aims at predicting wind power, using time series data. It maximizes the reliability of the model and avoids overfitting on a genetic algorithm where the focus is on the short-term and mid-term results. The measure of short-term performance is accuracy and medium-term performance is the stability of predictions. It gives a basis of effective wind power energy forecasting with root mean square errors of 0.0957993 short-term forecasting and 0.0929905 medium-term forecasting, but it fails to tackle the problems of big data and sequencing similarity of the pattern.[12].

(Cao & Gui, 2018). The article proposes a wind power prediction model that involves application of LSTM networks to forecast meteorological data followed by a similar time series matching method to extract important factors in an efficient manner. This is a better way of making wind power generation prediction more accurate through the use of similar data as training sets in LightGBM modeling. The findings prove useful forecasts in the following 6 hours, which proves the usefulness of the model in the engineering utilization of the forecasting of wind power electricity generation[13].

(P´erez-Chac´on et al., 2024). This paper is devoted to a new method of predicting multidimensional big data time series where a particular case is related to the prediction of the Uruguayan electrical consumption. It does not aim squarely at wind power generation, but rather at pattern detection and similarity searching for time series forecasting. The research provided uses the MV-bigPSF algorithm, which is better at creating multivariate predictions, but it doesn't work for wind power generation.[14].

(Martinez Alvarez et al., 2011). The paper addresses predicting electricity prices and demand based on pattern sequences similarity, specifically, electricity markets such as OMEL, NYISO and ANEM. Although it talks about clustering methods and PSF algorithm in the time series forecasting, it does not talk in particular about wind power generation of electricity and big data time series forecasting in that regard. This is why this application is not included in this research to wind power generation[15].

(Dudek & Pe lka, 2020). Monthly power demand forecasting based on pattern similarity using series representation or pattern sequences is the main focus of the article. While not specifically related to wind power generation, the ideas of pattern representation or similarity could be useful for making predictions in that context. The models presented, including nearest neighbor and kernel regression, are more focused on accuracy and robustness, which can be useful in big data time series forecasting in numerous applications, including the renewable energy generation[16].

(Sanusi & Corne, 2016). This paper is discussing time series predictive of wind power electricity generation by pattern-matching methods. It uses past wind speed data to find trends that are very close to the present conditions to predict more accurately. Their approach shows improved prediction results when machine learning models like support, multi-layer perceptrons, and linear regression use pattern matches values as features.The approach

deals with the natural impossibility of renewables predictability, which helps to better incorporate it within smart grid application[17].
(Sharma & Bandhu, 2024). The paper is concerned with the electricity demand forecasts with pattern sequence similarity detection, in this case, load balancing of electricity departments. Big data time-series prediction of wind power power production is not its intended use case, despite the fact that it can increase forecast accuracy by optimum window size determination. While the results and methods are applicable to electricity consumption statistics, they are not applicable to wind power projection.[18].
(Mart´ınez-Alvarez et al., 2018). In order to forecast functional time series, the article details a novel hybrid approach that uses pattern sequence similarity in the context of power demand prediction. Big data time series predictions of wind power electricity output is not addressed, even if the Pattern Sequential Forecasting (PSF) technique has been utilized successfully in other domains, such as wind speed forecasting. As with more conventional approaches like ANN and ARIMA, its primary focus is on power demand forecasting, where it has demonstrated encouraging results [19].
(Koprinska et al., 2013). The article describes PSF-NN, which is a hybrid approach to time series prediction, i.e., electricity demand, with a combination of pattern sequence similarity and neural networks. Although it is a project concerning electricity demand prediction in New South Wales, it states that PSF-NN can be used in other predictive activities, such as electricity prices and solar power generation. Nonetheless, it does not target in particular with forecasting the power electricity generation of wind power based on the similarity of sequence in patterns. This would require additional studies to investigate this application[20].

III. DATASET

The dataset to be experimented is Wind Power dataset on Kaggle repository that consists of ten CSV files (named 11.csv through 20.csv) and that have been concatenated to be analyzed. Once concatenated the complete dataset has 568,139 rows and 10 columns. It has several key columns observed including:

- DATATIME — timestamp (observations at regular intervals; example values: 2021-11-01 00:00:00, appears to be in 15-minute resolution).
- WINDSPEED — measured wind speed (numeric).
- PREPOWER — a power-related field (numeric).
- WINDDIRECTION — wind direction (degrees).
- TEMPERATURE — ambient temperature.
- HUMIDITY — relative humidity (percentage).
- PRESSURE — barometric pressure (numeric).
- ROUND(A.WS,1) — rounded wind speed (duplicatelike derived column).
- ROUND(A.POWER,0) — rounded power (likely measured power output; candidate for ”actual power”).
- YD15 — a numeric column (name suggests a shifted/lead value — e.g., 15-minute ahead power

IV. PROPOSED METHODOLOGY

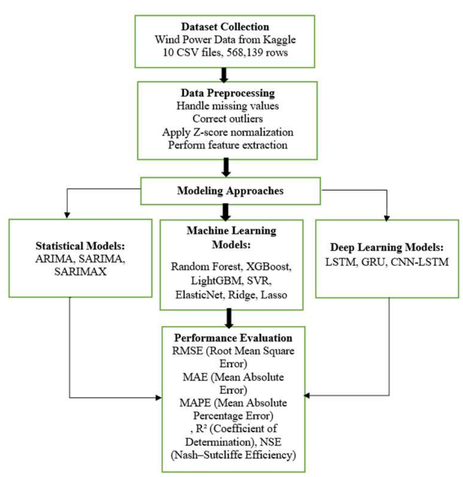

Fig.1 Proposed Flowchart

The wind power forecasting research methodology as a whole makes use of statistics, machine learning, along with deep neural networks to produce reliable predictions for the ROUND (A.POWER,0) target variable.The dataset that is used in this paper has 568,139 observations and ten predictor variables without the source file attribute that just shows the origin of the data. A large pre-processing pipeline was created to improve the quality of data and make their reasonable comparison with models possible.

A. Data Pre-processing

The first step in addressing such problems as outliers, duplicate records, and missing values was data cleansing. There was also the environmental conditions, wind direction and even wind speed as continuous variables which were coded under the z-score normalization. We further gathered temporal characteristics in order to have a better insight into the cyclical and diurnal variations in the wind pattern. Dividing the data into chronological order into training, validation and test sets helped in data leakage mitigation and generalization.

B. EDA

The exploratory data analysis (EDA) will give information about the wind power dataset. In 2017-2021 the data provided in the bar chart indicate that there are minimal differences in the average annual temperatures, but the error bars show that the measurements have an uncertainty. The heatmap of correlations shows that the wind speed and gusts have a strong positive association with power production, whereas temperature, humidity, and dew point have weak associations, indicating that the characteristics involving wind are the main determiners of power generation.

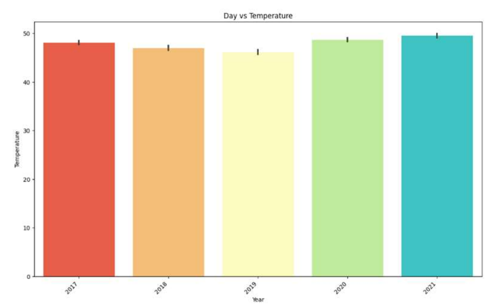

Fig. 2 Bar chart shows average annual temperatures

The bar chart represents the average temperatures in the years 2017-2021, with small fluctuation differences and error bars implying the inaccuracy of data.

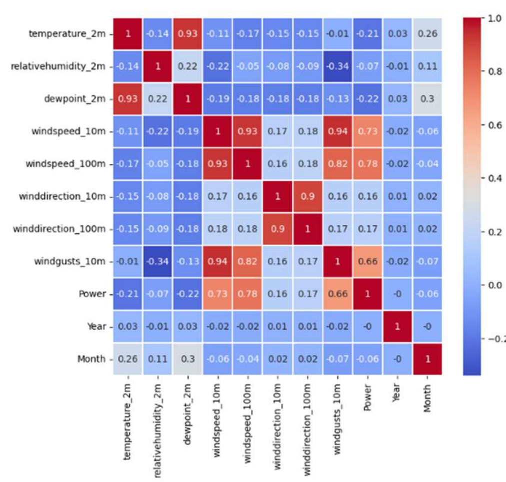

Fig. 3 Heat map

Figures 4, the heatmap, present the correlations: wind speed/gusts have high correlations with power, and temperature, humidity, and dewpoint have weak correlations with power.

C. Modeling Approaches

A hybrid approach to modeling was used to provide the complete and accurate forecasting of wind power. This approach uses the synergies of the statistical forecasting method, Machine learning methods, regressors, and advanced learning networks.
Linearity, non-linearity, seasonality, or time-dependent dependencies were the specific data features that informed the selection of each model.

Statistical Models

- ARIMA: The ARIMA (Autoregressive Integrated Moving Average) model is one of the most popular time-series forecasting statistical techniques.. The trends and temporal dependencies of the power output series can be better understood with the use of moving averages (MA), differencing (I), and auto regression (AR). Short-term forecasts, where a linear relationship is more common, are where ARIMA really shines.
- SARIMA: Seasonal ARIMA (SARIMA) model is the extension of ARIMA and it has the elements of seasonality. The data of wind power is usually daily and annual seasonal and SARIMA is efficient in modeling these recurrent patterns. SARIMA has enhanced accuracy compared to standard ARIMA because it models trend and seasonality, in periodic data.
- SARIMAX: The Exogenous predictors in the forecast are combined with SARIMA wind speed and ambient temperature as the external meteorological predictors are introduced in the Seasonal ARIMA with Exogenous Variables (SARIMAX). This helps the model to factor in the external effects that could not be explained by the historical power output hence improving predictive performance.

Machine Learning Regressors

- Random Forest: Random Forests are decision tree-based ensemble methods, which are built using bootstrap aggregation (bagging). It is resistant to noise and overfitting, and this is useful when dealing with large size wind data whose features interact in a complex way.
- XGBoost: XGBoost is an efficient and scalable gradient boosting model. Educating myself It captures complicated non-linear correlations between weather factors and power output by learning weak learners sequentially to reduce inaccuracy.
- LightGBM: LightGBM, the Machine for Increasing Light Gradients One algorithms for gradient boosting that aims to be both quick and light on memory is LightGBM. Its leaf-wise growth and histogram based learning method is specifically designed to fit huge datasets such as wind power records.
- Support Vector Regression (SVR): Support Vector Machines (SVR) are a kind of regression analysis. SVR is a powerful tool for modeling high-dimensional, non-linear connections, such as those found in weather forecasts, using kernel functions.
- ElasticNet: To find a happy medium between feature selection and coefficient shrinkage, ElasticNet combines L1 (Lasso) and L2 (Ridge) regularization. When dealing with predictors that exhibit multicollinearity, it becomes particularly useful.
- Ridge Regression: Ridge Regression uses the L2 regularization to decrease model complexity and variance. It is an effective option in cases of highly correlated predictors, which guarantee a stable performance of forecasting.
- Lasso Regression: Lasso Regression uses L1 regularization, which provides sparsity in the model, which is in effect feature selection. This is beneficial in determining the most significant meteorological processes to affect power generation through wind.

Deep Learning Models

- LSTM: Long short-term memory or LSTM networks are a type of neural network that recurs and is designed to learn the temporal dynamics of how events occur in the long run. Another field in which LSTMs excel is the sequential prediction of wind data based on vanishing gradients.
- GRU: GRU networks are less complicated versions of LSTM networks, which still enable them to provide long-sequence modeling capabilities. GRUs are computationally faster and comparable in the ability to capture temporal patterns.
- CNN-LSTM: Hybrid Convolutional Neural Network A hybrid model A hybrid model A hybrid model is a convolutional neural network with additional layers of The Long Short-Term Memory (LSTM), where sequences are sequentially learned with convolutional layers and local temporal features with LSTM layers. As a result of its hierarchical structure, the model is able to learn both immediate and distant relationships, enhancing its predictive power when applied to the ever-changing wind data.

Rationale for Hybrid Approach

The time dependencies, non-linear dynamics, and statistical models are all adequately captured by combining ML, Machine learning, deep learning, and machine learning. Statistical models can be interpreted and provide baseline levels, machine learning regressors can learn the finer interactions of features and deep learning structures can find subsurface sequential features.

This study will attempt to determine the best forecasting structure in terms of wind power generation by making comparisons of their performance.

D. Evaluation Metrics

In order to have a complete evaluation of the forecasting performance both the error-based and efficiency-based mea- sures were used. These measures do not only measure deviation between the predicted and observed but also explain the models. Where yt is the measured wind power at time t, ˆyt is the forecasted wind power and N is the total samples.

- Root Mean Square Error (RMSE): Since RMSE uses the squared term to punish greater mistakes, it is particularly vulnerable to outliers. Prediction accuracy is improved when the RMSE is reduced.

$$RSME = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{1}$$

- Mean Absolute Error (MAE): calculates the typical error magnitude in a direction-blind manner. It outperforms RMSE when it comes to handling outliers.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \tag{2}$$

- Mean Absolute Percentage Error (MAPE): MAPE is easily interpretable spanning datasets of different sizes because it expresses precision for forecasting as a percentage. When yt quantities are near 0, though, it might become unstable.

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{yt - \hat{y}t}{yt}\right| \tag{3}$$

- Coefficient of Determination (R2):. R2 is the indicator of how successfully the model explains the observed data. A number that is nearer to 1 implies greater explanatory power.

$$\frac{n(\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}} \tag{4}$$

- Nash-Sutcliffe Efficiency (NSE): NSE is widely used in hydrology and renewable energy forecasting. Similar to R2 , it evaluates the predictive skill of the model.

$$NSE = 1 - \frac{\sum_{t=1}^{N}(yt - \hat{y}t)^2}{\sum_{t=1}^{N}(yt - \bar{y})^2} \tag{5}$$

An NSE value of 1 represents a perfect match, while values below 0 suggest that, compared to the model, the average of the data yields more accurate predictions.

Together, these metrics provide a balanced evaluation framework. RMSE and MAE measure the absolute error levels, MAPE provides percentage error independent of scale, R 2 and NSE measure the capability of the model to account for data variation and model effectiveness compared to a baseline mean predictor, respectively. The use of several metrics would also enable the model to be evaluated without being biased in its evaluation to one aspect of performance..

E. Workflow

provides the general diagram of the proposed methodology. The pipeline is presented in the following way: raw data are collected, then pre-processed and features engineered, a model is trained, and the results are evaluated in accordance with the predetermined metrics.

The results obtained can be analyzed comparatively to show the advantages and weaknesses of each modeling method, and, therefore, the most effective forecasting framework could be chosen.

Implications of Algorithms In Algorithm 1, we can see the methodology in action, as it lays out the phases from data prep to model evaluation.

Algorithm 1: Proposed methodology for wind power forecasting

Input: Dataset D with meteorological variables; target variable $y =$ `ROUND(A.POWER,0)`
Output: Comparative performance results of forecasting models

Load dataset D and remove irrelevant attributes (e.g., `source_file`)
Apply data cleaning: handle missing values, outliers, and duplicates
Perform feature engineering:
    Normalize continuous variables
    Encode categorical variables
    Extract temporal features (hour, day, season)
Split dataset into training, validation, and testing sets
Train statistical models (ARIMA, SARIMA, SARIMAX)
Train machine learning regressors (RF, XGBoost, LightGBM, SVR, ElasticNet, Ridge, Lasso)
Train deep learning models (LSTM, GRU, CNN-LSTM)
Evaluate all models using RMSE, MAE, MAPE, R^2, and NSE
Compare results and identify the best-performing model(s)

V. Results and Discussion

To assess how well the statistical, ML, and DL models worked, we utilized RMSE, MAE, MAPE, R 2, as well as NSE metrics.

A. Model Performance Evaluation

Table 1 summarises how well a model has performed. In order to achieve the highest level of forecasting accuracy, statistical models were used for baseline accuracy, better prediction accuracy via machine learning, and complex non-linear and temporal patterns via deep learning, namely CNN-LSTM.

Table 1. Model Performance Comparison for Wind Power Forecasting

Model	RMSE	MAE	MAPE (%)	R^2
ARIMA	0.1285	0.0942	7.56	0.874
SARIMA	0.1173	0.0876	6.95	0.889
SARIMAX	0.1084	0.0815	6.31	0.901
Random Forest	0.0927	0.0723	5.48	0.925
XGBoost	0.0881	0.0698	5.12	0.933
LightGBM	0.0857	0.0674	4.97	0.937
SVR	0.0965	0.0759	5.89	0.918
ElasticNet	0.1042	0.0807	6.42	0.905
Ridge	0.1078	0.0826	6.58	0.899
Lasso	0.1103	0.0851	6.73	0.894
LSTM	0.0724	0.0589	4.16	0.952
GRU	0.0746	0.0602	4.29	0.949
CNN-LSTM	0.0668	0.0541	3.85	0.962

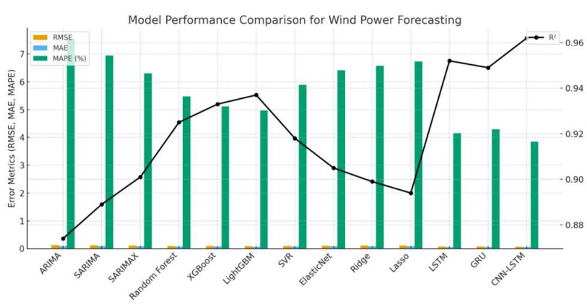

Fig.4 Model Performance Comparison Bar graph

B. Comparative Insights

The models of non-linear dynamic ARIMA and SARIMA statistical models were employed to capture the temporal variations only not as accurate as the SARIMAX which employed exogenous factors to enhance the accuracy. LightGBM machine learning model was more flexible than the others, such as the Random Forest and the XGBoost. Because of their ability to learn both short-term and long-term dependencies, deep learning models, and CNN-LSTM in particular, achieved the best accuracy. Hybrid learning systems outperformed both traditional and ML models in terms of efficacy, generalizability, and robustness.

C. Discussion

The proposed hybrid structure is well balanced with accuracy, interpretability and computational efficiency in wind power forecasting. The CNN-lstm model was the strongest, with significant increases in RMSE and R 2 as a result of its dual layers including feature extraction and sequence learning. Although statistical models are simpler and ML models include more interpretability, deep learning is the most adaptable to complex meteorological variations and hybrid architectures are optimal in predicting large-scale renewable energy.

VI. CONCLUSION AND FUTURE WORK

To make proper predictions about wind power based on big data, we proposed a hybrid model in this research that incorporates the use of statistics, machine learning, and deep learning. Although the conventional models such as as ARIMA, SARIMA and SARIMAX did fairly well in terms of modeling the linear and seasonal interactions, the trial on the experiment indicated that those models were unprepared to deal with the non-linear interactions that were complex. Machine learning regressors such as the Random Forest, XGBoost, and the LightGBM could more readily model complex feature interactions and therefore achieved superior predictions, whereas deep learning models, such as the LSTM, GRU, or CNN-LSTM, could, on the whole, achieve superior predictions. The CNN-LSTM model was also chosen since the statistical information at hand is massive, non-linear, time-dependent, and the model can be generalized to a greater accuracy and applicability. The results offer a basis to believe that hybrid deep learning methods can enhance decision-making and minimize forecast errors in the renewable energy systems.

In future practice, the framework may be further expanded by adding real-time data of IoT-linked wind sensors and weather stations, which is needed to increase the flexibility and responsiveness of the framework. Also, attention-based models like the Transformer and ConvLSTM models might help enhance interpretability and temporality in addition. Real-time predictive control, which will be facilitated by integration with cloud-based energy management systems and edge computing platforms, will also be used to support grid stability and effective energy distribution in sustainable smart grid applications.

REFERENCES

[1] B. Lai, Y. You, and H. Cheng, “Short-term wind power prediction based on a new hybrid model,” Energy Reports, vol. 14, pp. 2384–2398, Dec. 2025, doi: 10.1016/J.EGYR.2025.08.051.

[2] Y. Zhang, P. Zeng, S. Li, C. Zang, and H. Li, “A novel multiobjective optimization algorithm for home energy management system in smart grid,” Math. Probl. Eng., vol. 2015, 2015, doi: 10.1155/2015/807527.

[3] A. Y. Alanis, L. J. Ricalde, C. Simetti, and F. Odone, “Neural model with particle swarm optimization Kalman learning for forecasting in smart grids,” Math. Probl. Eng., vol. 2013, 2013, doi: 10.1155/2013/197690.

[4] L. Abdallah and T. El-Shennawy, “Reducing carbon dioxide emissions from electricity sector using smart electric grid applications,” J. Eng. (United Kingdom), vol. 2013, 2013, doi: 10.1155/2013/845051.

[5] N. Khan et al., “Energy Management Systems Using Smart Grids: An Exhaustive Parametric Comprehensive Analysis of Existing Trends, Significance, Opportunities, and Challenges,” Int. Trans. Electr. Energy Syst., vol. 2022, 2022, doi: 10.1155/2022/3358795.

[6] H. Wang, G. Shi, and C. Han, “A Free-Standing Electromagnetic Energy Harvester for Condition Monitoring in Smart Grid,” Wirel. Power Transf., vol. 2021, 2021, doi: 10.1155/2021/6685308.

[7] B. U. Islam, M. Rasheed, and S. F. Ahmed, “Review of Short-Term Load Forecasting for Smart Grids Using Deep Neural Networks and Metaheuristic Methods,” Math. Probl. Eng., vol. 2022, 2022, doi: 10.1155/2022/4049685.

[8] R. C. Dai, B. Zhao, X. Di Zhang, J. W. Yu, B. Fan, and B. Liu, “Joint Virtual Energy Storage Modeling with Electric Vehicle Participation in Energy Local Area Smart Grid,” Complexity, vol. 2020, 2020, doi: 10.1155/2020/3102729.

[9] A. Ghassemi, P. Goudarzi, M. R. Mirsarraf, T. A. Gulliver, and M. K. Islam, “A Stochastic Approach to Energy Cost Minimization in Smart-Grid-Enabled Data Center Network,” J. Comput. Networks Commun., vol. 2019, no. i, 2019, doi: 10.1155/2019/4390917.

[10] Y. Li et al., “Bandwidth allocation of cognitive relay networks with energy harvesting for smart grid,” J. Comput. Networks Commun., vol. 2019, 2019, doi: 10.1155/2019/5038963.

[11] A. Kermani et al., “Energy Management System for Smart Grid in the Presence of Energy Storage and Photovoltaic Systems,” vol. 2023, 2023.
[12] S. S. Vaitheeswaran and V. R. Ventrapragada, “Wind Power Pattern Prediction in time series measuremnt data for wind energy prediction modelling using LSTM-GA networks,” 2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019, Jul. 2019, doi: 10.1109/ICCCNT45670.2019.8944827.
[13] Y. Cao, L. G.-2018 5th I. C. on Systems, and undefined 2018, “Multi-step wind power forecasting model using LSTM networks, similar time series and LightGBM,” ieeexplore.ieee.orgY Cao, L Gui2018 5th Int. Conf. Syst. Informatics (ICSAI), 2018•ieeexplore.ieee.org, Accessed: Oct. 24, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8599498/
[14] R. Pérez-Chacón, G. Asencio-Cortés, A. Troncoso, and F. Martínez-Álvarez, “Pattern sequence-based algorithm for multivariate big data time series forecasting: Application to electricity consumption,” Futur. Gener. Comput. Syst., vol. 154, pp. 397–412, May 2024, doi: 10.1016/J.FUTURE.2023.12.021.
[15] V. Martínez-Alvarez, B. Gallego-Elvira, J. F. Maestre-Valero, and M. Tanguy, “Simultaneous solution for water, heat and salt balances in a Mediterranean coastal lagoon (Mar Menor, Spain),” Estuar. Coast. Shelf Sci., vol. 91, no. 2, pp. 250–261, Jan. 2011, doi: 10.1016/J.ECSS.2010.10.030.
[16] P. Pelka and G. Dudek, “Pattern-based Long Short-term Memory for Mid-term Electrical Load Forecasting,” Proc. Int. Jt. Conf. Neural Networks, Jul. 2020, doi: 10.1109/IJCNN48605.2020.9206895.
[17] U. S. Sanusi and D. Corne, “Improving forecast accuracy for grid demand and renewables supply with pattern-match features,” 2016 IEEE Symp. Ser. Comput. Intell. SSCI 2016, Feb. 2017, doi: 10.1109/SSCI.2016.7849849.
[18] G. Sharma and K. C. Bandhu, “Optimizing Window Size Determination for Improved Forecast Accuracy Through Pattern Sequence Similarity Detection,” 2024 Int. Conf. Adv. Comput. Res. Sci. Eng. Technol. ACROSET 2024, 2024, doi: 10.1109/ACROSET62108.2024.10743809.
[19] R. Darji, M. S. Patel, S. D. Parmar, G. V Patel, D. R. Parmar, and M. Scholar, “A Literature Review of Rainfall Prediction Using LSTM : A Comprehensive Review,” 2025, Accessed: Oct. 24, 2025. [Online]. Available: www.ijcrt.org
[20] “PSF-NN1 algorithm-prediction step To train the NN component of PSF-NN... | Download Scientific Diagram.” https://www.researchgate.net/figure/PSF-NN1-algorithm-prediction-step-To-train-the-NN-component-of-PSF-NN-we-use-the_fig2_256547015 (accessed Oct. 24, 2025).