International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 3, Issue 2, 135-146, 2022
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V3I2P114

Original Article

AI-Driven Load Forecasting for Smart Grids under High Renewable Penetration

Krishna Gandhi[1], Pankaj Verma[2]
[1]Illinois State University, 100 N University St, Normal, IL 61761, United States
[2]Indian Institute of Management, Bangalore (IIM-Bangalore), Bannerghatta Road, Bengaluru, Karnataka, India.

Abstract *- The introduction of renewable energy sources into the contemporary power system also brings a lot of variability, uncertainty, and nonlinearity in electricity demand, which makes the conventional load forecasting techniques challenging. Artificial intelligence (AI) and machine learning (ML) have become promising methods to capture complex temporal and spatial relationship of load data, which provide better prediction capabilities of renewable-integrated smart grids. This paper critically evaluates AI-based load predictive designs, which comprise artificial neural networks, support networks, decision tree ensembles, and deep learning devices, e.g., recurrent neural networks, long short-term memory networks, and gated recurrent units and convolutional neural networks. Hybrid and ensemble models, which combine statistical models and signal decomposition methods, are also presented. The main variables affecting the forecasting performance, such as the meteorological, socio-economic, and grid-level variables are mentioned, as well as the data sources, feature engineering approaches, and metrics. Issues of data quality, generalization of models, scalability, and uncertainty are discussed and future research possibilities, such as probabilistic forecasting, real-time adaptive learning, and integrated load-renewable modeling are mentioned. This review will offer a unified source of information on how AI-based load forecasting practices can be developed and implemented within a renewable-abundant smart grid setting to researchers and practitioners.*

Keywords *- Smart Grids, Load Forecasting, Artificial Intelligence, Machine Learning, Renewable Energy Integration, Deep Learning, Hybrid Models, Ensemble Learning, Uncertainty Modeling.*

1. Introduction

A major transformation is being experienced in the electric power systems due to the growing demand of electricity, the necessity to ensure greater reliability and development of new and better communication and control technology. Old power grids, mainly developed to be centralized and with a simple pattern of demand, are increasingly being transformed into smart grids that can accommodate the possibility of a two-way (bidirectional) power flow, distributed energy sources, and smart decision making processes. The objective of this transition is to improve the efficiency of the operations, the resilience of the systems, and the good utilization of the available energy resources.

One of the most important features of the contemporary power system is an increasing concentration of renewable energy sources, e.g. wind and solar power. These are resources that are part of sustainable energy production, however they are also new challenges as they are not only variability, but also less controllable. Digitalization, the development of improved sensing infrastructure, and the use of analytical data based on data has facilitated the transition in the conventional grids to smart grids and then to renewable-based smart grids as seen in Figure 1. In this regard, artificial intelligence has come out as a prospective device to deal with the complexity that comes with the operation of modern power systems.

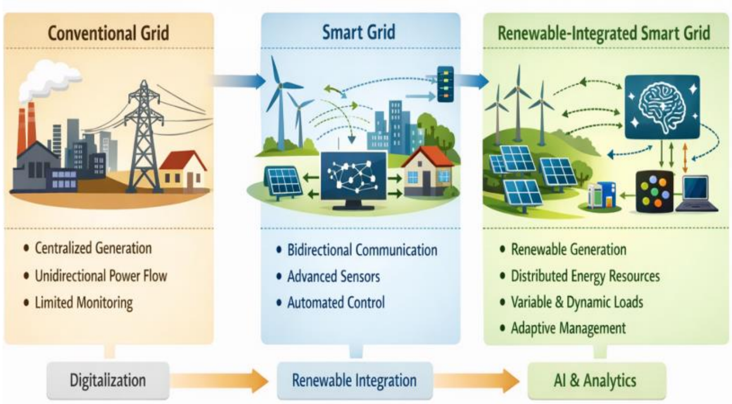

Fig 1: Evolution of Power Systems from Conventional Grids to AI-Enabled Smart Grids

Figure 1 illustrates the structural evolution of electrical power systems from conventional grids toward intelligent and renewable-integrated smart grids. Conventional grids are characterized by centralized power generation, unidirectional power flow, and limited monitoring capabilities. The transition to smart grids introduces advanced sensing infrastructure, bidirectional communication, and automated control mechanisms that enable real-time monitoring and improved operational efficiency. With the increasing integration of renewable energy sources, smart grids further evolve into renewable-integrated systems that exhibit higher complexity due to distributed generation, variable power output, and dynamic load behavior. Artificial intelligence and data-driven analytics play a central role in this evolution by enabling adaptive decision-making, enhanced load forecasting, and efficient management of uncertainty in modern power systems.

Proper modeling and forecasting of demand of electricity has therefore gained a growing importance. Predictive load performance under dynamic operating conditions is also paramount to stability of the system, efficient resource distribution and provision of stable power supply under renewable-rich operating conditions.

Renewable energy sources are highly penetrated and this has a great impact on the nature of operation of the power system. Contrary to the traditional generation units, wind and solar power is intermittent and unpredictable with a high dependency on weather patterns. The variability presents nonlinear interaction between the generation and demand, which makes it challenging to balance the provider and the consumer.

This is due to the growing application of changing renewable resources, which cause rapid changes in net load curves, peak shifting, and less predictability of system behavior. Such impacts may create discrepancies on the generation and demand of electricity, thus creating more pressure on grid operations. System operators therefore need sophisticated forecasting instruments that are capable of modeling sophisticated time trends and unpredictability related to integrations of renewable energy.

The load forecasting is an important aspect of smart grid planning and operation. Precise predictions of loads are used in diverse operations, such as unit commitment, economic dispatch, reserve allocation and congestion management. Moreover, good predictions allow practicing the demand response plan efficiently and integration of distributed energy resources.

Load forecasting, in renewable-integrated smart grids, does not simply focus on what demands are going to be used, but also what happens between the demand and the variable generation. The accuracy of the forecasts leads to a better grid stability, lower operation expenses, and the increased use of renewable energy, and as a result contributes to the sustainable functioning of the power system.

The traditional load forecasting methods, which include statistical regression models and autoregressive time-series methods, have been used extensively because they are simple and easy to understand. These techniques generally assume some form of linearity and stationarity of the load patterns which restricts their applicability to complex and dynamic environments.

In highly renewable penetrative states, demand has a strong nonlinear pattern that is dependent on various situations such as weather variability, distributed generation, and consumer actions. Conventional forecasting models tend to fail in the ability to model these complex relationships, thus leading to lower prediction accuracy and lower adaptability to shifting operating conditions.

The review is an extensive discussion of load forecasting methods based on artificial intelligence utilized in smart grids that have large integration of renewable energy. It focuses on popular AI and machine learning strategies that have been widely developed in current literature. The review includes short-term, medium-term and long-term horizons of load forecasting with the emphasis on the way they are applied and the methodological issues.

Moreover, this work reviews the benefits and drawbacks of various AI-based forecasting models, addresses the issue of hybrid and ensemble approaches, and highlights the major issues that are connected with the quality of data, its generalization, and uncertainty management. Through the synthesis of the participating research works, the review is expected to provide insightful information to the researcher and practitioners involved in the load forecasting in renewable-integrated smart grid settings.

2. Load Forecasting In Smart Grids: Fundamentals

2.1. Definitions and Forecasting Horizons

Load forecasting is defined as the process of estimating the future electricity demand in a given time frame and it is a background element of power system planning and operation. Load forecasting is divided into short-term, medium-term and long-term forecasting, depending on the prediction horizon and the application needs. Both categories vary according to temporal resolution, factors of variation, and purpose.

Short-Term Load Forecasting (STLF) is generally concerned with the demand in electricity on the basis of minutes or some few days. STLF is mainly applied in the real time grid operation, unit commitment, economic dispatch, and frequency regulation. The correct short-term predictions are essential towards ensuring that the systems are stable and operational costs kept at the lowest.

Medium-Term Load Forecasting (MTLF) deals with the forecast of maintenance schedules, fuel purchases, and resource planning that are from one to several weeks up to months. MTLF needs to capture the seasonal patterns, weather trends and consumption patterns.

Long-term load forecasting (LTLF) takes into account long-term horizons, i.e. more than several months to several years and assists in strategic planning activities, e.g. generation expansion, transmission investment, policy analysis. LTLF focuses on long-term trends of demand growth and socio-economic factors. Table 1 presents different forecasting horizons and their application.

Table 1: Load Forecasting Horizons and Applications

Forecasting Horizon	Time Scale	Primary Applications	Key Influencing Factors
Short-Term (STLF)	Minutes to days	Unit commitment, economic dispatch, grid stability	Weather conditions, historical load, time
Medium-Term (MTLF)	Weeks to months	Maintenance planning, fuel scheduling, resource allocation	Seasonal trends, demand patterns
Long-Term (LTLF)	Months to years	Capacity expansion, infrastructure planning	Economic growth, population, policy

2.2. Load Characteristics under Renewable Integration

The assimilation of renewable sources of energy changes the nature of the electricity demand that is monitored by the power system greatly. The more predictable and smooth traditional load profiles are being substituted by unpredictable, and dynamic net-load profiles. Unpredictable shifts in renewable generation increase demand volatility through a rapid alteration in the system load levels.

Shifting of peaks is one of the most noticeable features of renewable-integrated systems as the peak net demand is reached at atypical hours because the renewable output is also not constant. The difference between the total electricity demand and the renewable generation is referred to as the net load, which is especially crucial in the given context. The high penetration of solar power is usually followed by low net loads during the day and high ramping loads during the evening, a phenomenon also known as the duck curve.

The changing nature of these load characteristics also causes further complexity to the forecasting models in that they need proper modeling to capture the nonlinear interactions of the demand, generation, and external factors.

2.3. Key Factors Affecting Load Forecasting Accuracy

Forecasting the correct loads requires proper integration of the various factors that contribute to changes in the consumption of electricity. Such factors can be generalized into meteorological, socio-economic and behavioral, as well as grid-level operational factors.

2.3.1. Meteorological Factors

Among the most powerful drivers of electricity demand are the meteorological conditions which are important in short term forecasting of electricity demand. The heating and cooling loads directly depend on temperature, and thermal comfort and air-conditioning consumption depend on humidity. The velocity of the wind and the level of sun rays directly affect electricity usage and creation of renewables, indirectly influencing the behavior of net-loads. The non-linear dependence of weather variables on electricity demand implies the advanced modeling methods that are able to address the complex dependencies.

2.3.2. Socio-Economic and Behavioral Factors

The socio-economic and behavior variables contribute highly towards the formation of long and medium-term electricity demand. Urbanization and the population growth are factors which affect the threshold rates of electricity usage. The shift in consumer behavior such as the introduction of electric vehicles and the use of appliances that are electrically powered also play a part in transforming the nature of demand. Such considerations present slow yet continuous changes in load profiles that have to be considered in forecasting models.

2.3.3. Operational Factors and grid-level.

Operation mechanisms at grid levels also influence the electricity demand and predictability as well. The demand response program changes the consumption by motivating users to change or decrease demand during peak hours. Generations that are decentralized (like solar installations on rooftops) will decrease the net demand to the grid and create variability. The storage systems of energy also alter the load behavior by allowing the shift in the consumption of electricity over time. All these operational factors together make the system more complex and pose other difficulties to the precise load forecasting.

3. Traditional Load Forecasting Approaches

The conventional load forecasting techniques were founded on the principles of statistics and have been extensively used in the analysis of the power system owing to the mathematical clarity and simplicity of its implementation. The methods mainly depend on historical loads information and assumption on how the data will behave. They tend to work well under fairly stable operating conditions but may have poor performance in complex and dynamic environments that are highly variable and interact nonlinearly.

3.1. Time-Series Statistical Methods

Time-series statistical models are models used to predict future load values based on temporal dependence of historical demand data. The methods presuppose that historical load patterns are rich in information to forecast future behavior, and are used when the data has regular and repetitive structures.

3.1.1. Autoregressive Models (AR, ARMA, ARIMA)

Autoregressive-based models describe the electricity demand as linear sum of the past values of the electricity demand plus stochastic error. Autoregressive (AR) models are used to model the dependence between load observations at a certain point in time and the ones before it, whereas Autoregressive Moving Average (ARMA) models use both the autoregressive and moving averages. ARIMA models are built on this framework by including the operation of differencing to solve non-stationary data.

Although they are highly used, the models are based on the assumption of stationarity and linearity, which limit them in modeling non-linear load dynamics. The autoregressive models are less effective in renewable-integrated power systems, where electricity demand tends to have nonlinear trends, affected both by the variable production and external factors.

3.1.2. Techniques of exponential smoothing.

Exponential smoothing techniques also make the prediction of the load with the use of weighted averages of the previous values whereby the most recent values carry more weight. The holt winters gets the basic exponential smoothing but includes trend and seasonal and can therefore be used appropriately to model periodic demand changes.

The models are seasonally adjusted through the use of seasonal adjustment methods, which allow them to represent the repetitive trends in the consumption of electricity. Nevertheless, exponential smoothing procedures presuppose constant cycling patterns and slow changes in trends. These assumptions do not necessarily hold under high variability and high load changes as the forecasting can be limited.

3.2. Regression-Based Models

Forecasting models in regression determine explicit relationships between electricity demand and variables of influence. Linear regression models are used to predict the load based on the explanatory variables, including temperature, humidity and the calendar effects. The framework can be further developed using multiple regression methods which add more predictors to enhance the explanatory power.

The models are appreciated due to their interpretability and ease of implementation. Nonetheless, regression-based approaches normally tend to suppose that there exist linear relationships between inputs and output. Physically, load demand and external factors are highly nonlinear and context-dependent especially in systems where there is distributed renewable generation and demand-side participation.

3.2.1. Limitations in Renewable-Rich Environments

Even though the traditional forecasting models proved to be effective in the stable power systems, their drawbacks become even more evident in the areas that are rich in renewables. The growing range and ambiguity of the renewable power sources present a challenge to the fundamental principle of linearity, stationarity, and predictable seasonal variation of these models.

Conventional techniques find it difficult to respond to the dynamic nature of net-load profiles, shifting of peaks and multifaceted interactions between demand and variable generation. Moreover, they have low capacity to include high dimensional input features which limit their performance when several factors that influence the outcome are to be adjusted at the same time. Table 2 presents the main weaknesses of traditional load forecasting methods in power systems that include renewable integration.

Table 2: Limitations of traditional forecasting models under renewable penetration

Model Type	Key Assumptions	Major Limitations in Renewable-Rich Systems
AR / ARMA / ARIMA	Linearity, stationarity	Poor handling of nonlinear and volatile load patterns
Exponential Smoothing	Stable trends and seasonality	Inability to adapt to rapid demand fluctuations
Linear Regression	Linear input–output relationship	Limited representation of complex dependencies
Multiple Regression	Fixed predictor relationships	Reduced scalability with increasing feature complexity

4. AI Techniques for Load Forecasting

4.1. Overview of Artificial Intelligence in Power Systems

The application of artificial intelligence (AI) methods to power system operation and planning has been on the increase because of the ability to represent nonlinear relationships that are difficult to model and adjust to changing conditions. Originally, the first systems that were introduced were expert systems to encode the rules of operation and aid in decision-making. As computational resources and data availability improved, machine learning (ML) methods became popular as a tool to do predictive modeling, where systems learn patterns using past data without having to be explicitly programmed. In recent times, deep learning (DL) algorithms have been used to predict the existence of complex time-varying and spatial relationships in load data, enhancing the accuracy of forecasts when operating under conditions of variable loads. Figure 2 offers a taxonomy of commonly used AI methods in load forecasting.

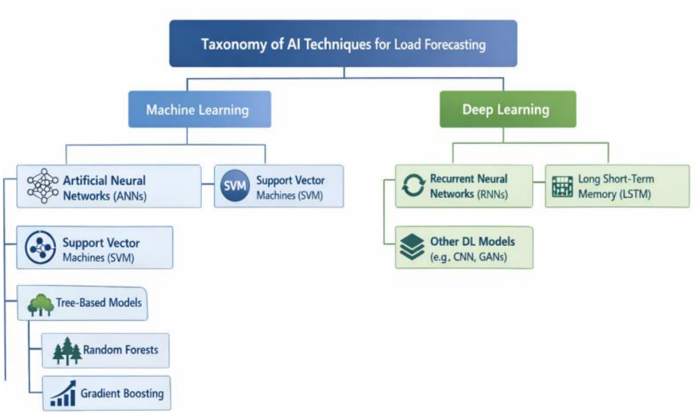

Fig 2: Taxonomy of AI Techniques for Load Forecasting

Figure 2 illustrates the hierarchical classification of AI methods applied to electricity load forecasting. At the top level, AI techniques are divided into machine learning and deep learning approaches. Machine learning methods include artificial neural networks (ANNs), support vector machines (SVM), and tree-based models such as random forests and gradient boosting machines. Deep learning methods, which are a subset of neural networks, are capable of capturing complex temporal and spatial patterns in load and renewable generation data. This taxonomy highlights the diversity of AI approaches and their respective suitability for handling nonlinearities, high-dimensional input features, and variable renewable penetration in modern power systems.

4.2. Artificial Neural Networks (ANNs)

Artificial neural networks have also been used extensively in load forecasting because they can approximate nonlinear associations amid the input characteristics and the electricity demand. They are composed of interdependent wiring of fake neurons which process input information via weighed connections and activation functions.

4.2.1. Neural networks: Feedforward Neural Net.

One of the earliest uses of ANN in load forecasting (short term) was in feedforward neural networks (FNNs). Data flows in these networks without feedback loops with the input layer fed to the hidden layers and then to the output layer. FNNs are able to obtain nonlinear relationships between load and causative factors like temperature, humidity and calendar effects. Initial research showed that the use of FNNs led to substantial gains in terms of accuracy compared to the use of conventional statistical models.

4.2.2. Backpropagation-Based Learning.

The training of feedforward networks is normally based on training using the backpropagation algorithm, which changes the network weights to reduce prediction error. Backpropagation-based learning, though effective, may be faced with challenges like overfitting where the network architecture is complicated with respect to the available data. Commonly used methods to improve generalization performance include regularization methods, early stopping and cross-validation.

4.3. Support Vector Machines (SVM)

Support vector machines are the supervised learning algorithms which have the capability of performing regression and classification jobs by identifying the best hyperplanes within feature spaces, which are of high dimensions. SVMs are resistant to overfitting and they can take nonlinear relationships by using the kernel functions.

4.3.1. Support Vector Regression (SVR)

The SVM principles are applied to regression problems with the objective of approximating target values within a given tolerance by Support vector Regression. RBF and polynomial kernels are some examples of kernel functions that help SVR to express nonlinear relationships between input variables and electricity demand. It is demonstrated that SVR works well in the case when the input features are high-dimensional and the load patterns are complex.

4.3.2. Applications in Renewable-Aware Forecasting

Application SVR has been extensively used in load prediction systems with a large proportion of renewable. SVR models can be used to forecast the net-load patterns more accurately than the linear statistical models because it uses weather variables, like the wind speed, solar irradiance, and temperature. They are used in short-term wind and solar-integrated load forecasting and hybrid models that have used SVR with decomposition models.

4.4. Decision Trees and Ensemble Learning

Decision tree–based models and ensemble learning methods offer alternative AI approaches that are interpretable and capable of handling nonlinear relationships.

4.4.1. Random Forests

Random forests are made out of a collection of decision trees that are built on various subsets of the training data and features. The method minimizes variation and increases predictability. The importance of features can be measured using random forests and thus the most significant factors on the load demand can be identified. They have shown good capability in complex load dynamics capture under renewable integration.

4.4.2. Gradient Boosting Machines

Gradient boosting machines (GBMs) recursively create an ensemble of weak learners which are usually shallow decision trees voting on prediction residuals by optimization. GBMs enhance precision in short-term forecasting load assignments, and it is able to model nonlinear relationships among load and various external factors. They are tailored well to systems with distributed renewable generation, because of their iterative nature and high-dimensional capabilities.

4.5. Comparison of Significant AI Techniques.

Table 3 provides a comparative introduction to commonly used AI methods to forecast the load with their strengths and limitations highlighted.

Table 3: Comparison of Major AI Techniques Used for Load Forecasting

AI Technique	Key Strengths	Limitations / Challenges	Typical Applications
Feedforward Neural Networks	Nonlinear modeling, flexible architecture	Overfitting, requires tuning, data-intensive	Short-term load forecasting
Backpropagation Learning	Effective weight optimization	Sensitive to initialization, local minima	STLF, MTLF
Support Vector Regression	Handles high-dimensional inputs, robust	Kernel selection critical, slower training	Renewable-aware load forecasting
Random Forests	Nonlinear modeling, feature importance	May require many trees, less interpretable	STLF, feature selection
Gradient Boosting Machines	High accuracy, handles complex dependencies	Risk of overfitting, computationally intensive	Short- and medium-term forecasting

5. Deep Learning Approaches

The predictability of load has been a growing focus in deep learning because of its capacity to capture complicated temporal and spatial interactions among the data on electricity demand. In contrast to classical AI systems, deep learning models have the capability to learn high-level features of raw input signals as well as learn long-term sequential variations, which are especially appropriate with renewable-integrated smart grids.

5.1. Recurrent Neural Networks (RNN)
Recurrent neural networks are specially created to deal with time-dependent data by including feed-back connections to enable the transfer of information through time. With this architecture, RNNs are able to derive temporal information in load data, which is essential to both short-term and medium-term forecasting. Nonetheless, in the traditional RNNs, there may be challenges in learning long term dependencies because of the vanishing or exploding gradient issues, which restrict their capability to model long sequences of time.

5.1.1. Long Short-Term Memory (LSTM)
Long Short-Term Memory networks are a higher form of RNN which attempts to fix the shortcomings of the traditional RNNs in that they explicitly address long-term dependencies with memory cells and gating mechanisms.

5.1.2. LSTM Architecture
The LSTM networks are composed of memory cells that are governed by three main gates namely the input gate, forget gate and the output gate. These gates control the exchange of information and ensure that the network is sensitive enough to hold any information of relevancy in history and the unnecessary information are eliminated. This also lends LSTM specific applications to load forecasting problems where the usual demand of electricity is determined by the short-term and long-term historical data. LSTMs have proved to be more accurate in the modeling of nonlinear and variable patterns of loads than the conventional RNNs.

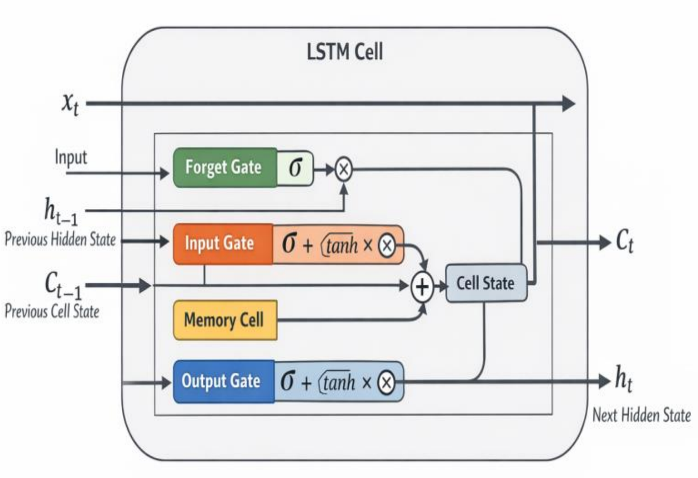

Fig 3: LSTM Cell Structure for Load Forecasting

Figure 3 depicts the internal architecture of an LSTM cell used for load forecasting. The memory cell maintains a state vector that stores relevant temporal information, while the input, forget, and output gates control the flow of data into and out of the cell. The input gate regulates the incorporation of new information, the forget gate determines which information to discard from the memory, and the output gate decides the information passed to the next time step. This gated structure allows LSTM networks to capture long-term dependencies in sequential load data, making them highly effective for forecasting electricity demand in systems with high variability.

5.2. Gated Recurrent Units (GRU)
Gated Recurrent Units This is a simpler form of LSTM that uses the forget gate and the input gate together as an update gate. This design is less complex to compute and has not lost the ability to represent temporal dependencies in sequential data. GRUs are especially effective in problems where computational performance is a bottleneck, yet still able to give similar forecasting performance as LSTMs in most load prediction problems.

5.3. CNN-Based Load Forecasting
Full CNNs have been modified to load forecasting with the consideration of the input maps of the spatial and temporal patterns. As an example, CNN layers can be used to extract features by passed to load maps describing consumption per region or weather grids including meteorological variables. CNN-based methods can be used to integrate both space and time associations to improve forecasting accuracy in multi-faceted systems of renewable integration, as both methods combine convolutional layers with temporal modeling.

6. Hybrid and Ensemble Forecasting Models
To address the drawbacks of the use of single statistical or AI-based models, hybrid and ensemble forecasting models have been presented. The hybrid models by integrating various techniques can take the advantages of both techniques to enhance the

confidence and accuracy of predictions, especially in the renewable-integrated smart grids with the very high variability and nonlinearity in load patterns.

6.1. AI–Statistical Hybrid Models

Statistical hybrid models combine classic time-series methods with artificial intelligence methods to improve the quality of forecasts. Examples are common ARIMA-ANN and Wavelet-ANN. The ARIMA-ANN models have the advantage that ARIMA part will catch linear trends and seasonality, whereas the ANN part will replicate the nonlinear patterns of residual values leading to a better overall forecast precision. Equally, Wavelet-ANN models initially reduce the load signal to various frequency bands via wavelet transform and subsequently use ANN to forecast the signal of each frequency band. Such hybrid models have been extensively used in short-term and medium-term load prediction showing better results when compared to independent models.

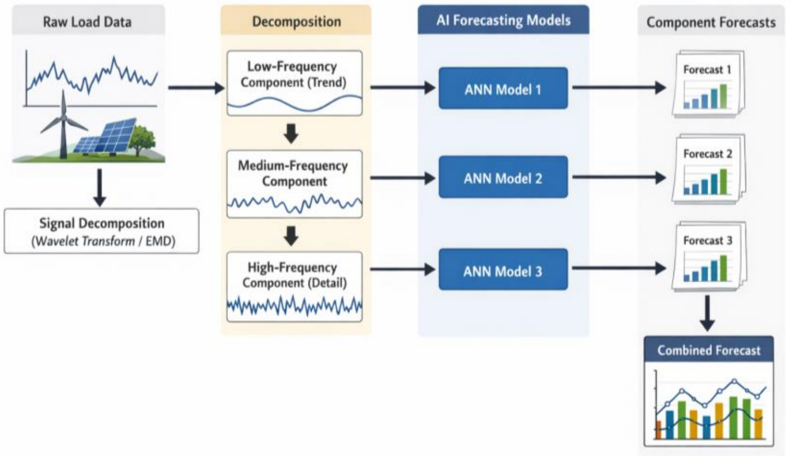

Fig 4: Hybrid Load Forecasting Framework Combining Decomposition and AI

Figure 4 illustrates a hybrid load forecasting architecture where raw load data is first decomposed using signal processing techniques such as wavelet transform or empirical mode decomposition. Each decomposed component is then processed by an AI model, such as an artificial neural network, to generate individual forecasts. The component forecasts are subsequently combined to produce the final load prediction. This hybrid approach allows effective modeling of both linear and nonlinear patterns in load data, while also capturing high-frequency fluctuations introduced by renewable generation.

6.2. Signal Decomposition Techniques

Hybrid models can also employ signal decomposition methods to decompose load signals into various components that reflect various frequency or trend attributes. This preprocessing step can enhance the performance of forecasting and result in a smaller model since it removes complexity and the capacity of AI models to identify pertinent patterns.

6.2.1. Wavelet Transform

Wavelet transform breaks down a time series into approximate and detailed components in a number of scales. This will enable it to have long-term trends extracted and short-term seasonal changes in the load data. Hybrid frameworks can more accurately model the slow seasonal trends and those that change rapidly due to the presence of renewable generation by using AI models on individual wavelet components.

6.2.2. Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition is a data-driven method that breaks down a non-stationary signal into intrinsic mode functions (IMFs) representing oscillatory modes. Similar to wavelet-based decomposition, EMD can improve the prediction of complex load patterns when combined with AI models. Each IMF is modeled independently and aggregated to produce the final forecast, reducing error due to high-frequency noise or nonlinear interactions.

6.3. Ensemble Learning Strategies

Ensemble learning is a technique that unites several prediction models so as to increase accuracy and minimize the uncertainty in predictions. The simple approaches are model averaging, where forecasts of various models are averaged to get the final generation. In weighted ensembles, each of the model predictions is weighted by its performance measure and the more accurate the model, the higher the weight affecting other models.

Ensemble techniques may involve statistical, AI, and deep learning models used together, which enables them to be resilient to nonlinearities, seasonality, and volatility brought about by renewable generation. They work well especially in systems where uncertainty is high and the individual models might not explain all the dynamics of electricity demand.

Table 4: Overview of Hybrid and Ensemble Forecasting Models

Model Type	Key Features	Advantages	Typical Applications
ARIMA-ANN	Linear ARIMA + nonlinear ANN	Captures both linear and nonlinear patterns	STLF, MTLF
Wavelet-ANN	Signal decomposition + ANN	Handles multi-scale load variations	Renewable-aware forecasting
EMD-AI	EMD decomposition + AI models	Reduces high-frequency noise, improves accuracy	STLF, MTLF
Model Averaging Ensemble	Multiple model forecasts averaged	Simple, robust to outliers	STLF, MTLF
Weighted Ensemble	Weighted combination of forecasts	Prioritizes accurate models, reduces variance	Renewable-integrated load forecasting

7. Data, Features, and Evaluation Metrics

Proper load forecasting is not only dependent on the modeling method used but also on the quality of the data as well as the selection of pertinent features and the application of the right measures of performance evaluation. This part gives an overview of the popular data sources, feature engineering methods, and evaluation measures used in AI-based load forecasting of smart grids.

7.1. Data used to make a load forecast

Load forecasting models are usually based on more than one source of data which complement each other. Smart meters produce high-resolution household or facility-level consumption data, and in that way, it allows forecasting short-term and medium-term predictions. SCADA systems provide operational information like the loadings for the substations, like the generator output, and the health of the network, which are essential in system-wide prediction. Weather related variables like temperature, humidity, wind speed, and solar irradiance metrics are made available in meteorological stations and are major determinants of electricity demand, especially in a system where there is a high penetration of renewable energy.

7.2. Feature Engineering Methods

One of the processes that should be implemented to enhance the accuracy of forecasting is feature engineering which converts raw data to informative predictors that reflect the corresponding patterns. Popular techniques are the modeling of temporal dependencies using lagged values of loads, calendar effects, weekdays, weekends, and holidays to model a regular pattern of consumption, weather normalization to deal with environmental change. The combination of the two enables AI models to capture more accurately short-term changes as well as long-term trends in electricity demand.

Table 5: Common input features used in AI-based load forecasting

Feature Category	Example Features	Purpose / Contribution
Historical Load	Previous hour/day/week load	Capture temporal dependencies and trends
Calendar Effects	Day of week, holidays, seasonal indicators	Model routine patterns and seasonality
Meteorological Variables	Temperature, humidity, wind speed, solar irradiance	Represent environmental influence on load
Derived Features	Moving averages, load differences, lagged features	Enhance model sensitivity to short-term fluctuations

7.3. Performance Evaluation Metrics

Some quantitative measures are typically employed in order to determine the accuracy and reliability of load forecasting models. Mean Absolute Error (MAE) is used to gauge the magnitude of errors in average without taking their direction into account, it is a simple measure of prediction accuracy. Root Mean Square Error (RMSE) puts more emphasis on large errors, which emphasizes extreme deviations. Mean Absolute Percentage Error (MAPE) standardizes the errors against the actual values and makes sense to compare the systems with different levels of load. The Normalized Root Mean Square Error (NRMSE) once again scales the RMSE by the range of observed data to give a dimensionless value that can be used to make benchmarks.

These metrics of evaluation make it possible to compare various AI and hybrid forecasting models systematically and select models to apply in practice to the implementation of renewable-based smart grids with their integration.

8. Challenges, Open Issues, and Anticipated Future Directions

Although there has been a major improvement in the field of AI-based load forecasting in smart grids, there are various challenges that have limited the feasibility of these models. The quality and accessibility of the data are important issues since the absence of measurements, sensor noise, and anomalies in smart meter and SCADA data may compromise the quality of forecasting. Also, privacy concerns limit the availability of finer consumptions data, which is challenging to the training and validation of models.

Other major limitations are model generalization and scalability. Most AI and hybrid forecasting models are region-specifically tuned and feature selected and hence lack transferability to other geographical locations or system configurations. This restricts the generalizability of the models that have been formulated to a specific grid or load profile. There is a constant challenge of ensuring that forecasting models, which can easily make decisions in the changing conditions of the system but which do not compromise accuracy.

The forecasting of the demand and electricity generation is also complicated by uncertainty in electricity demand and renewable production. Quantile regression and ensemble-based predictive techniques are early probabilistic forecasting techniques that have been designed to measure prediction uncertainty. The approaches yield useful information that can be used to make risk-conscious decisions, but they are usually constrained by the complexity of the calculations and the presence of adequate historical data.

In the future, a number of avenues of research are expected to drive the area of load forecasting in smart grids that are incorporated with renewable energy. It is anticipated that by integrating the load and renewable generation forecasting, predictive performance will be improved as both demand and variable renewable will be modeled. Amplified focus is expected on probabilistic and risk-conscious forecasting methods that will offer more informative forecasting to the system operators and will take uncertainty and extreme events into account. Moreover, real-time and adaptive learning systems are expected to gain more in the future so that models can be constantly updated and adapted to new circumstances, such as new load patterns and variability of renewable generation. By solving such challenges and investigating such research directions, it is likely to enhance the accuracy, reliability, and operational decision-making in contemporary smart grids.

9. Conclusion

The conversion of conventional power systems into renewable-combined smart grids has essentially altered the outlook of load forecasting. The large penetration of the variable renewable energy sources like wind and solar can cause uncertainty and intermittency in the electricity demand and non-linear dynamics that make the traditional forecasting techniques inadequate. Conventional statistical models, such as autoregressive and regression-based models, are interpretable and easy to understand, but they are unable to fully characterize the dynamics of modern power systems. They can only be used with limited success in renewable rich environments because of their assumptions of linearity, stationarity and predictable seasonal behavior, leading to poorer accuracy and adaptability.

Machine learning and artificial intelligence methods have become one of the potent alternatives, as they are able to represent nonlinear dependencies, process high-dimensional input features, and respond to evolving operating conditions. Neural networks Feedforward neural networks, support vector regression, decision tree ensembles, and gradient boosting machines have shown substantial gains in the accuracy of their predictions when compared to the traditional methods. The deep learning architectures, such as recurrent neural networks, long short-term memory networks, gated recurrent units, and convolutional neural networks, provide the capacity to learn long-term temporal dependencies and spatial correlations, and improve the performance, especially in the short and medium-term forecasting.

Hybrid and ensemble models, which are AI mixed with statistical models or signal decomposition methods, e.g. wavelet transform and empirical mode decomposition, offer extra robustness and accuracy by leveraging the relative advantages of each particular model. They are especially effective at dealing with high-frequency variations, nonlinear interactions, and multi-scale variations added by renewable generation to allow more robust and reliable load forecasts.

The success of AI-based forecasting models relies heavily on the quality of available data, the chosen wisely input features, and the application of the right metrics of performance evaluation. The electricity demand is affected by meteorological, socio-economic, behavioral, and grid-scale operational aspects and an adequate integration of those factors in the forecasting model is necessary to reflect the intricate dynamics of the contemporary power systems. Standardized measures like MAE, RMSE, MAPE, and NRMSE allow assessing the predictive performance across models and applications and make them comparable.

In spite of the improvements, there are still a number of issues. Unreliable and inaccurate model results may be caused by data quality and availability, such as missing measurements, noise, and privacy restrictions. The issues of model generalization and scalability are also of significant concern, since most AI and hybrid models need to be fine-tuned to a specific region to be more transferable to other grids. Also, the uncertainty that is inherent in renewable production and demand leads to the need to do probabilistic or risk-aware forecasting, which, though potentially useful, is computationally expensive and data-intensive.

The future research paths that may be expected involve incorporation of load and renewable generation forecasting that have the potential of enhancing predictive performance through the joint modeling of demand and variable renewable generation. More focus is likely to be on probabilistic and risk-aware forecasting techniques that can deliver more information to system operators, including uncertainty and extreme events. Real-time and adaptive learning systems will be developed, which will allow forecasting models to respond to changes in the conditions of the system, the emergence of new patterns of loads, and the transformation of renewable generation profiles. These adaptive systems are bound to improve the performance of the operations, the reliability of the grids and the use of renewable energy.

To sum up, AI-powered load prediction is an innovative breakthrough in operations of the smarter grid, and it may enable handling the nuances of the resource-rich contemporary power infrastructure, which is highly penetrated with renewables. Remedying current issues and following the expected research trends, the future forecasting models will be able to be more accurate, reliable, and adaptable, thus making sustainable, resilient and economically efficient operations of the power systems. The findings in this review are intended to inform the researcher and practitioners to come up with effective AI-based load prediction methods that will satisfy the dynamic demands of renewable-integrated smart grids.

References

[1] H. I. Henderson, D. Novosel, and M. L. Crow, "Electric power grid modernization trends, challenges, and opportunities," IEEE Power Energy, 2017.
[2] D. Infield and L. Freris, Renewable energy in power systems. John Wiley & Sons, 2020.
[3] I. Alotaibi et al., "A comprehensive review of recent advances in smart grids: A sustainable future with renewable energy resources," Energies, vol. 13, no. 23, p. 6269, 2020.
[4] S. M. Amin, "Smart grid: Overview, issues and opportunities. advances and challenges in sensing, modeling, simulation, optimization and control," European Journal of Control, vol. 17, no. 5-6, pp. 547–567, 2011.
[5] S. Sun, M. Dong, and B. Liang, "Distributed real-time power balancing in renewable-integrated power grids with storage and flexible loads," IEEE Transactions on Smart Grid, vol. 7, no. 5, pp. 2337–2349, 2015.
[6] S. Impram, S. V. Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: A survey," Energy Strategy Reviews, vol. 31, p. 100539, 2020.
[7] I. J. Perez-Arriaga and C. Batlle, "Impacts of intermittent renewables on electricity generation system operation," Economics of Energy & Environmental Policy, vol. 1, no. 2, pp. 3–18, 2012.
[8] J. Oyekale et al., "Impacts of renewable energy resources on effectiveness of grid-integrated systems: Succinct review of current challenges and potential solution strategies," Energies, vol. 13, no. 18, p. 4856, 2020.
[9] S. R. Khuntia, J. L. Rueda, and M. A. M. van Der Meijden, "Forecasting the load of electrical power systems in mid-and long-term horizons: a review," IET Generation, Transmission & Distribution, vol. 10, no. 16, pp. 3971–3980, 2016.
[10] E. Aguilar Madrid and N. Antonio, "Short-term electricity load forecasting with machine learning," Information, vol. 12, no. 2, p. 50, 2021.
[11] X. Liang, "Emerging power quality challenges due to integration of renewable energy sources," IEEE Transactions on Industry Applications, vol. 53, no. 2, pp. 855–866, 2016.
[12] S. C. Sekhar, "Thermal comfort in air-conditioned buildings in hot and humid climates–why are we not getting it right?," Indoor Air, vol. 26, no. 1, pp. 138–152, 2016.
[13] A. K. Singh et al., "Load forecasting techniques and methodologies: A review," in 2012 2nd International Conference on Power, Control and Embedded Systems, IEEE, 2012.
[14] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir," Journal of Hydrology, vol. 476, pp. 433–441, 2013.
[15] J. Ponoćko, Data analytics-based demand profiling and advanced demand side management for flexible operation of sustainable power networks. Springer Nature, 2020.
[16] I. K. Nti et al., "Electricity load forecasting: a systematic review," Journal of Electrical Systems and Information Technology, vol. 7, no. 1, p. 13, 2020.
[17] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," Analyst, vol. 135, no. 2, pp. 230–267, 2010.
[18] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," Applied Energy, vol. 238, pp. 1312–1326, 2019.
[19] M. S. H. Lipu et al., "Artificial intelligence based hybrid forecasting approaches for wind power generation: Progress, challenges and prospects," IEEE Access, vol. 9, pp. 102460–102489, 2021.
[20] M. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," IEEE Access, vol. 7, pp. 125413–125423, 2019.

[21] A. Al Mamun et al., "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," IEEE Access, vol. 8, pp. 134911–134939, 2020.
[22] O. Tshenyego, R. Samikannu, and B. Mtengi, "Wide area monitoring, protection, and control application in islanding detection for grid integrated distributed generation: A review," Measurement and Control, vol. 54, no. 5–6, pp. 585–617, 2021.