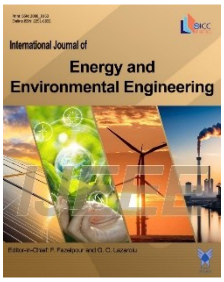

Volume 15, Issue 4, (2024) Article 16

International Journal
of
Energy and Environmental Engineering (IJEEE)

https://doi.org/10.57647/ijeee.2024.1504.16

A Systematic Review of Data Analytics Approaches in Smart Electricity Markets

Muhammad Muwahid Asim[1], Fareed Ud Din[2,*], Asif Karim[3], Niusha Shafiabady[4,5]

[1] *Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute, Topi, Pakistan*
[2] *School of Science and Technology, University of New England, Armidale, NSW, Australia*
[3] *Faculty of Science and Technology, Charles Darwin University, Casuarina NT 0909, Australia*
[4] *Department of IT, Australian Catholic University, North Sydney, New South Wales, Australia*
[5] *Faculty of Science and Technology, Charles Darwin University, Haymarket, New South Wales, Australia*

***Corresponding author:** fuddin@une.edu.au

Review Article

Abstract

The rapid digitalization, decentralization, and integration of renewable energy in electricity markets have heightened the demand for advanced data analytics to manage their complexity and ensure sustainability. This systematic review explores the transformative role of data analytics, machine learning and artificial intelligence in enhancing efficiency across key areas of smart electricity markets. These include forecasting (price, demand, load, carbon emissions, and supply), trading, security, bidding, and the impact of renewable energy on pricing. Following PRISMA guidelines to conduct this systematic literature review (SLR), a rigorous selection process narrowed 504 initial articles to 23 high-impact studies from 2014 to 2024, offering a focused analysis of emerging techniques such as hybrid models, neural networks, blockchain, and explainable AI. The review identifies advancements in predictive accuracy, adaptive decision-making, and decentralized trading mechanisms, emphasizing their ability to address market volatility and enhance operational stability. Furthermore, the findings highlight gaps in existing approaches and propose future research directions, including the integration of real-time analytics, reinforcement learning, and decentralized frameworks. By providing a structured overview, this study serves as a resource for researchers and practitioners, underscoring the critical role of data-driven solutions in the evolving landscape of smart electricity markets.

Keywords: Smart Electricity Market, Renewable Energy, Data Analytics, Artificial Intelligence, Machine Learning, Explainable AI

Cite this article: Asim M.M., Din F.U., Karim A., Shafiabady N., A Systematic Review of Data Analytics Approaches in Smart Electricity Markets. Int. J. Energy Environ. Eng. 2024; 15(4) : 1-20 https://doi.org/10.57647/ijeee.2024.1504.16

1. Introduction

Smart electricity markets are undergoing rapid transformation driven by digitalization, decarbonization, decentralization, and the increasing penetration of renewable energy sources [1]. These changes have significantly increased the complexity of power systems, creating a critical need for advanced data analytics to ensure operational stability, forecasting accuracy, and efficient market performance. The ongoing digitalization, decarbonization and decentralization of the smart electricity market have made the role of data analytics more crucial. Recent advancements in computing power, combined with the declining cost of information and communication technologies, have led to greater data availability and new opportunities for advanced analysis [2]. Furthermore, the growing volatility of electricity generation has increased complexity and thus creates a new need for data analytics [3]. At the same time, global electricity marketization is continuously evolving due to which deregulation, introduction of competition and commoditization have become a general trend [4]. Recent research highlights a paradigm shift from traditional statistical models to advanced machine learning and deep learning techniques and hybrid architectures. These approaches have been successfully applied to key areas of smart electricity markets such as forecasting, bidding, trading, and security, offering significant improvements in predictive accuracy and adaptability. Emerging technologies such as Explainable AI (XAI) and blockchain are further advancing transparency, interpretability, and real-time decision-making capabilities.

Within this context, the applications for data analytics in smart electricity markets includes forecasting [3], trading [4], bidding [5], renewable energy impact on price [6] and security [7]. Collectively, these studies provide valuable insights into distinct subfields of data analytics and highlight the need for a comprehensive review. Accordingly, this systematic review examines these studies across the various domains, identifies innovative approaches being introduced, and compares them with traditional methodologies.

Forecasting represents one of the most prominent application areas and can be further classified into price forecasting [8], demand forecasting [9], carbon emission forecasting [10], supply forecasting [11] and load forecasting [12]. Accurate forecasting is essential for effective planning and the safe operation of power systems [13]. However, forecasting is often challenged by the influence of external factors such as seasonal variations, weather patterns, electricity price fluctuations, and holidays [11]. Due to the inherent non-linearity in these relationships, machine learning models demonstrate superior capability in capturing complex patterns and achieving higher accuracy compared to traditional statistical methods [13]. Moreover, hybrid models are increasingly being adopted to enhance performance by integrating the strengths of multiple approaches [9].

In addition to these improvements, a diverse range of novel approaches has emerged, including Neural Networks [8], blockchain technology [6], Explainable AI [15], and fuzzy logic [16]. These techniques are applied across different domains of the electricity market, showcasing their versatility and potential for innovation. For instance, Azadeh Kheirandish et al. [16] utilized fuzzy logic to describe and analyze the behavior of fuel cell electric bicycle systems, further illustrating how such advanced methods are contributing to the broader landscape of smart electricity analytics. These papers provide valuable information on the specific sub-fields of data analytics in the smart electricity market. In this systematic review, we aim to look into all these papers across different fields of data analytics in smart electricity markets, the new innovative approaches being introduced, and their comparison with existing traditional approaches.

While previous studies have examined individual aspects such as price forecasting or market security, there remains a lack of a comprehensive review that systematically analyzes and compares the diverse data analytics approaches applied across multiple domains of smart electricity markets. Addressing this gap, this study consolidates findings across forecasting, trading, bidding, renewable energy pricing, and security, presenting a structured synthesis of state-of-the-art methodologies, their advantages over traditional methods, and emerging research trends.

2. Research Methodology

This review rigorously follows the methodology presented by Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [17]. To choose papers for the systematic review, we performed a series of steps starting with choosing databases for research articles, followed by deciding on keywords for the search in the databases. This was followed by removing duplicates and then removing articles based on journal ranking. Moreover, after this step, articles were filtered based on the abstract to get the most relevant articles. Lastly, from the pool of these filtered papers, we performed one final round of filtering based on the full text of the articles. This gave a final list of articles that were most relevant to Data Analytics in the Smart Electricity Market, and these articles were finally used for the systematic reviews. Figure 1 represents the PRISMA flow diagram.

2.1. Database and Keywords

The first phase of selecting papers was choosing databases for the research articles. We choose IEEE Xplore, Scopus and Springer databases for searching research articles for Data Analytics in the Smart Electricity Market. The next phase was deciding keywords and filters for the databases. The keywords used for the databases were TITLE-ABS-KEY ((("data analytic") OR (“AI”) OR (“classif”) OR ("machine learning") OR ("artificial intelligence") OR (“forecast”)

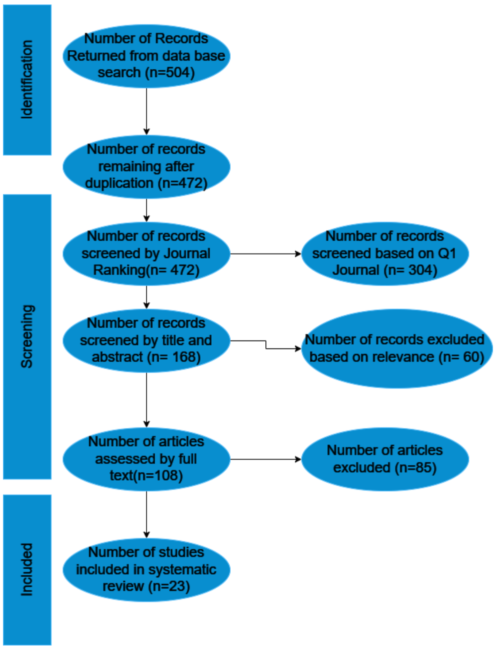

Figure 1. PRISMA Flow diagram

OR (“predict”))) AND (("energy market") OR ("electricity market") OR ("smart electricity market"))). The filter was set for paper in the last 10 years, from 2014 to 2024. This gave a total of 239 research articles from Scopus, 236 research articles from IEEE Xplore and 29 research articles from Springer. This gave a total of 504 research articles.

2.2. Duplication and Rank-based Filtering

After having a pool of 504 research articles, the first phase of filtering consisted of removing duplicate research articles. There were a total of 32 duplicate articles, reducing the number from 504 to 472. The next step was to filter all research articles that were not published in Q1 journals to get a pool of high-impact research articles. After this phase, we had a total of 168 research articles left. These 168 research articles were further filtered in the next step.

2.3. Abstract and Full-Text Filtering

A total pool of 168 research articles was left after filtering based on journal ranking. The next phase of filtering was to remove irrelevant research articles after reading titles and abstracts. This way we can get a pool of relevant research articles. In this phase, we filtered a further 60 research articles. There were a total of 108 relevant research articles left. Lastly, these 108 research articles were to be filtered based on the full text of the research articles. This way we can get research articles that align perfectly with the systematic review topic and with the needs of the systematic review. After this full-text filtering, the 23 most relevant papers were chosen to be analyzed in depth for the systematic review. In these 23 research articles, we analyzed the field of application, the approach being used and its comparison with traditional methods. Figure 2 shows a pie chart to represent the distribution of selected papers based on publication year.

2.4. Thematic Analysis

The most commonly used measures for comparison of approaches are Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). RMSE measures the average difference between predicted values and the actual values. The lower the value of the Root Mean Squared Error, the better the model is. The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. MAPE is a measure of the prediction accuracy of a forecasting method in statistics. RMSE, MAE and MAPE are given in equations (1), (2), and (3).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\Sigma(y_i - \hat{y}_i)} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\Sigma|y_i - \hat{y}_i| \tag{2}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\Sigma\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

In this systematic review, we aim to identify the most commonly used data analytics techniques in analyzing smart electricity markets, and how have data analytics methods been applied to forecasting electricity price, demand and supply. Additionally, this study aims to present the current and emerging trends in data analytics techniques and potential future advancements in the smart electricity market and how will they be beneficial. This review thematically structures the findings as per specific domains, which also represents different sections in the paper according to the application of data analytics in the smart electricity markets. Table 1 represents the fields of applications of the papers explored and the approach used. Future research should explore advanced optimization techniques. These techniques include metaheuristics (e.g., genetic algorithms, particle swarm optimization) to automate and optimize model architecture and hyper-parameter selection.

Table 1. Selected Papers, the field of applications and approaches used

Paper	Application	Approach
Z Qu et al. [8]	Short-term price forecasting	Improved wavelet neural network (IWNN)
X Fang et al. [9]	Short-term price forecasting	A hybrid model combining CatBoost for feature selection and BDLSTM for forecasting
C. Zhang and Y. Fu [18]	Probabilistic Price Forecasting	Extreme Learning Machine integrated with optimal prediction intervals
Cixin Xiao et al. [19]	Pre-dispatch price forecasting	Online Sequential Extreme Learning Machine
Wei Sai et al. [20]	Wholesale Electricity Price Forecast	XG-Boost algorithm
A. T Eseye et al. [21]	Short-term demand forecasting	Hybrid models for feature selection combining Binary Genetic Algorithm (BGA) and Gaussian Process Regression (GPR)
Mehedi et al. [12]	Short-term demand forecasting	Intelligent machine learning with an evolutionary algorithm-based Short-Term Load Forecasting (IMLEA-STLF)
Alex Coronati et al. [22]	Residual demand Curve forecasting	Hybrid LSTM approach in combination with one-dimensional and convolutional LSTM networks
Choujun Zhan et al. [23]	Short-term Load forecasting	Generative and Moving Interactive Neural Networks (GMINN)
Qin Shen et al. [24]	Short-term Load forecasting	Multi-Scale Ensemble Method (MSEM)
Yi Xuan et al. [25]	Short-term Load forecasting	A hybrid neural network combining a Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (BiGRU)
M. Rafiei et al. [26]	Probabilistic Load Forecasting	Combining Generalized Extreme Learning Machine (GELM), Improved Wavelet Neural Networks (IWNN)
M. Rostam Niakan Kalhpri et al. [27]	Long-term Hourly Load forecast	SVM, RF, and ANN in each step with fuzzy logic in the inference engine
Vahid Aryai et al. [10]	Carbon Emission Forecasting	Particle Swarm Optimized-extremely randomized trees (PSO-ERT) regression model
Zehang Li et al. [28]	Supply Forecasting	Hierarchical clustering method based on a weighted distance tailored to non-bounded and non-smooth curves
Mizue Shimomura et al. [6]	Impact of Renewable energy on price	Machine Learning and Explainable AI
Wenxuan Liu et al. [4]	Trading behavioral Modeling	Hybrid Experimental Learning (HEL)
Dhaou Said [29]	Decentralized trading framework	Combining consortium block-chain technology, Machine Learning (ML), and game theoretic models
Y.Li et al. [30]	Virtual Bidding	Machine Learning
Dong Han et al. [31]	Virtual Bidding	Conditional value-at-risk (CVaR) and deep reinforcement learning (DRL)
Yueyong Yang et al. [32]	Bidding Recommendation	Selective learning
Yikun Huang et al. [33]	False Data Injection	Deep Belief Networks (DBNs)
Peyman Razmi et al. [34]	Collusion Detection	Classification and regression tree (CART) and support vector machine (SVM)

This has the potential to improve model performance and scalability across diverse smart electricity markets. Moreover, there is a need to enhance decision-making under uncertainty and moving towards probabilistic methodologies will be beneficial in this regard. Techniques such as quantile regression and probabilistic graphical models can provide valuable insights into the range of possible outcomes, crucial for risk management and operational planning.

3. Forecasting in the Smart Electricity Market

Forecasting plays an important role in managing and integrating energy resources in the smart electricity markets. Accurate forecasts are crucial for maintaining grid stability and balancing energy resources efficiently. The variable nature of solar and wind is also addressed with the help of forecasting. The precision of predictions are improved by utilizing data analytics, advanced machine learning and statistical models. Forecasting is categorized into price forecasting, demand forecasting, load forecasting, carbon emission forecasting and supply forecasting.

3.1. Price Forecasting

Predicting accurate electricity price forecasting (EPF) is an important element in the domain on smart electricity markets [8]. However, this task is complicated by several challenges, including inherent volatility, non-stationarity, and multi-seasonality of electricity prices, which often reduce the effectiveness of traditional forecasting models. To address these challenges, the authors proposed a novel hybrid model for day-ahead EPF, which integrates feature extraction, pattern recognition, neural networks, and machine learning techniques. The approach, known as an Improved Wavelet Neural Network (IWNN), is initialized using an Extreme Learning Machine (ELM) to enhance prediction accuracy and convergence speed. Moreover, historical daily electricity price curves are categorized into distinct patterns through Lasso Regression, k-mean clustering, and Support Vector Machine (SVM) for pattern recognition.

Furthermore, EPF models play a crucial role for power generators, helping them optimize operational schedules, and for distributed energy resource (DER) aggregators, enabling them to design optimal bidding and scheduling strategies [8]. Alternate, models such as Autoregressive Integrated Moving Averages (ARIMA) [35] exist; however, they often failed to deliver the desired accuracy due to their limited ability to capture market impacts on electricity prices.

In contrast, machine learning (ML) models, such as Support Vector Regression (SVR), and deep learning (DL) methods—including Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Wavelet Neural Networks (WNN) – have demonstrated superior capabilities in handling nonlinear time-series data, making them more effective alternatives.

Among these, WNNs combine wavelet transform localization and neural network self-learning capabilities, offering high prediction accuracy and fast convergence. However, traditional WNNs have issues such as local optima and slow or non-convergence due to unsuitable initial parameter values and the exponential growth of training samples. To overcome these limitations, the IWNN model, supported by ELM initialization, optimizes initial network parameters, enhancing training speed and prediction accuracy while avoiding overfitting. Figure 3 below shows the RMSE of the proposed IWNN approach and other approaches including WCP, RNN and ARIMA.

The IWNN model has shown better performance than state-of-the-art forecasting algorithms, achieving lower error metrics in empirical studies. We can observe that IWNN with RMSE 1.9852 outperforms other approaches with the next best being RNN with an RMSE of 4.4592 which is an improvement of 55.7% as compared to RNN. Its ability to handle the complexities of day-ahead electricity price forecasting makes it an ideal choice for market participants.

Deep neural networks (DNNs) have not been extensively studied or applied to electricity price forecasting within specific markets such as Nord Pool [36]. Advanced DNN architectures and hybrid models that could enhance forecasting accuracy are therefore explored. A novel hybrid model combining the categorical boosting (CatBoost) algorithm for feature selection and a bidirectional long short-term memory (BDLSTM) network as the forecasting engine.

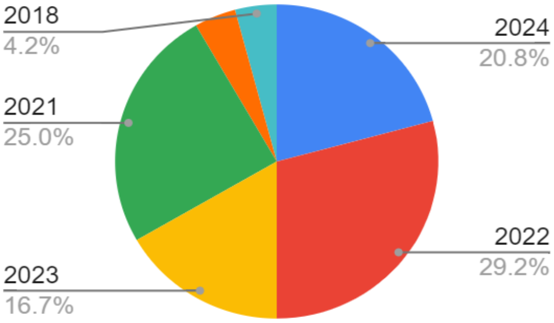

Figure 2. Year-wise distribution of selected paper

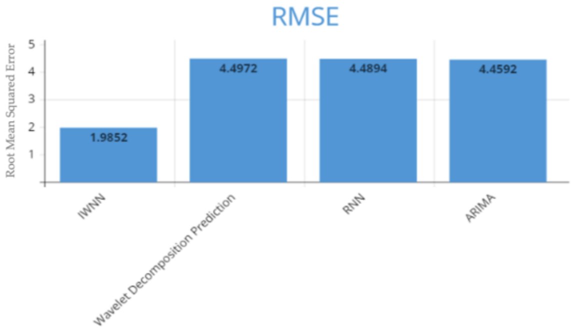

Figure 3. RMSE for IWNN and other approaches for price forecasting

CatBoost automates the conversion of categorical variables into numerical representations. This reduces preprocessing efforts and mitigates overfitting through a random permutation mechanism during dataset division. The BDLSTM network addresses the gradient vanishing problem. It also captures dependencies within the time series more effectively by preserving information from both past and future data points. The model was validated using 2018 hourly electricity price data from the Nord Pool market. Comparative performance analysis with other models, including multi-layer perceptron (MLP), support vector regression (SVR), ensemble trees, ARIMA, and recent deep learning models like GRU and LSTM, revealed that the proposed CatBoost-BDLSTM hybrid model achieved lower forecasting errors (measured by MAPE, RMSE, and MAE). Figure 4 shows the RMSE for different approaches. The proposed model shows RMSE 43.013 outperforming other approaches with the next best approach being LSTM with an RMSE of 45.807. Despite the higher computational cost in terms of training and forecasting time, the model's superior accuracy makes it a valuable tool for stakeholders in the electricity market.

On the other hand, deregulated electricity markets have a complex landscape [37], where accurate price forecasting is more important for all market participants. Traditional electricity price forecasting can be divided into point forecasts and probabilistic forecasts. Point forecasts provide a single predicted value for each time point. This offers straightforward interpretations but fails to account for uncertainties inherent in price predictions. Probabilistic forecasts offer a range of potential outcomes. Hence, addressing these uncertainties, this novel probabilistic electricity price forecasting method that emphasizes the creation of optimal prediction intervals (PIs). These intervals consider both reliability ensuring the forecast accuracy and sharpness, ensuring narrow interval widths, which is critical for decision-making confidence. The core of this method involves using quantile regression to estimate the upper and lower bounds of the PIs without making assumptions about the price distribution. This non-parametric approach avoids the common pitfalls of assuming normal or logistic distributions, which often do not fit the skewed nature of electricity prices. Extreme Learning Machine (ELM) is integrated to enhance the model's capability to capture the nonlinear relationships within electricity price data. The effectiveness of this probabilistic forecasting method is validated using data from various eletricity markets. Comparative analyses demonstrate that this method outperforms traditional models e.g., autoregressive integrated moving average (ARIMA). It also outperforms neural networks combined with bootstrap strategies. This method provides a robust tool for market participants. This is achieved by avoiding the assumption of price distribution and effectively capturing nonlinear relationships in the data. This method also holds promise for forecasting not just electricity prices but also power system loads and renewable energy generation, ensuring more confident and efficient system operations.

Similarly, pre-dispatch price forecasting is another critical component in the smart electricity market [19]. Traditional offline batch-learning methods usually fall short in responding promptly to sudden changes in the local power system environment. They are influenced by dynamic price fluctuations in neighboring regions. To address these limitations, a novel approach utilizing an online sequential extreme learning machine (OS-ELM) algorithm is proposed. This method is validated through extensive simulations using data from the Australian electricity market. This demonstrates improved accuracy in forecasting under unexpected conditions. Conventional methods often rely on linear models such as ARIMA [35]. They may not effectively capture the nonstationary nature of electricity prices. Advanced techniques like artificial neural networks (ANNs) and deep learning models offer better performance. They still suffer from limitations due to their dependence on batch-learning processes. These processes are time-consuming and require extensive training data which makes them less responsive to real-time changes. To overcome these challenges, OS-ELM, a member of the ELM family, is introduced that can perform real-time data training. The OS-ELM algorithm is capable of continuous updating. This makes it more adaptive to unexpected environmental changes. This adaptability is crucial given the increasing penetration of intermittent distributed energy resources and the growing complexity of modern smart grids [14]. The OS-ELM approach leverages a 2-D orthogonal list data structure, enabling it to process and update training patterns for both daily and 2-hourly data series.

This dual-module system comprises OS-ELM-1, which provides day-ahead predictions, and OS-ELM-2, which offers rolling 30-minute forecasts.

The first OS-ELM in this approach module updates its training pattern daily. This provides a rolling day-ahead price prediction with quantified prediction intervals (PI). The second module updates every two hours. This offers dynamic 30-minute forecasts that incorporate recent price information from neighboring grids. This integration helps capture potential price spikes and enhances overall forecasting accuracy. The novelty of this approach is in its hybrid structure and the implementation of a 2-D orthogonal list to increase prediction accuracy. Simulation results from the Australian electricity market underscore the method's effectiveness in improving forecast accuracy, particularly in scenarios involving unexpected changes in local and neighboring markets. The Proposed online learning forecasting gives an RMSE of 0.068 as compared to 0.1455 for batch learning which proved to be the next best approach based on RMSE which shows an improvement of 53.26%. This advancement in pre-dispatch price forecasting marks a significant step forward in the application of data analytics in the smart electricity market.

The real-time event-driven forecast model for predicting wholesale electricity is another application of smart electricity markets similar such as present by Wei Sai et al. [20] in Singapore's liberalized energy market using the XGBoost algorithm. In Singapore, maintaining a stable grid necessitates a significant reserve margin to mitigate the risk of outages or disruptions. However, reliance on spinning reserve units entails inefficiencies and increased costs. Accurate price forecasting is crucial for market participants to adjust generation and consumption strategies effectively, particularly in the context of increasing renewable energy penetration. The event-driven forecast model represents a significant advancement in energy price forecasting. This offers improved accuracy and responsiveness in a dynamic electricity market.

3.2. Demand Forecasting

Accurate demand forecasting including short-term demand forecasting and residual demand curves forecasting is important for efficient operation and planning within decentralized electricity markets [34]. Traditional forecasting models often fail to capture the unique patterns of smaller decentralization. To address this gap, recent advancements have focused on integrating machine learning techniques for enhanced feature selection and demand prediction such as the work presented by A. T. Eseye et al. [21] to improve short-term electricity demand forecasting in decentralized energy systems through integrated feature selection using machine learning techniques. Feature selection becomes crucial to enhance prediction accuracy while reducing computational complexity and overfitting risks. This hybrid feature selection method combining the Binary Genetic Algorithm (BGA) and Gaussian Process Regression (GPR). BGA is employed for feature selection. It leverages its robust searching ability and suitability for high-dimensional problems. GPR serves as the fitness measure for evaluating selected features, capturing nonlinear input-output relationships and providing probabilistic estimates of prediction uncertainty. The selected feature subsets outperform alternative feature selection methods, leading to significant accuracy improvements in electricity demand forecasting.

The electricity demand forecasting model developed using the obtained FS results has achieved an annual accuracy improvement of 38.7%, 81.2%, 81.9% and 83.0%, respectively for residential, educational, office and mixed-used building types compared to forecasting based on the original feature space without FS as shown in Table 2. The findings underscore the importance of integrating effective feature selection techniques with forecasting models to achieve robust and accurate predictions.

Conventional forecasting methods relied on manual techniques and basic statistical tools. Recent advancements in artificial intelligence (AI) and machine learning (ML) have revolutionized load prediction, offering more accurate and efficient models such as intelligent machine learning with an evolutionary algorithm-based Short-Term Load Forecasting (IMLEA-STLF) model for power systems [12]. IMLEA-STLF model integrates wavelet transform (WT) for data decomposition, oppositional artificial fish swarm optimization algorithm (OAFSA) for feature selection, and water wave optimization (WWO) with Elman neural networks (ENN) for prediction. This holistic approach aims to enhance prediction accuracy.

The novelty of the IMLEA-STLF model lies in its comprehensive methodology, encompassing data preprocessing, feature selection, prediction, and parameter tuning stages.

Table 2. Annual Accuracy Improvement for demand forecasting

Category	Accuracy
Residential	38.7
Educational	81.2
Office	81.9
Mixed Use Building Types	83.0

The IMLEA-STLF model's accuracy in forecasting power system loads outperforms existing methods in terms of predictive accuracy. By employing multi-level decomposition and feature selection techniques, the model achieves higher precision in load forecasting, offering valuable insights for power system operators and stakeholders. Figure 5 shows the MAPE (%) of proposed IMLEA-STLF and other approaches. The MAPE (%) of the proposed IMLEA-SLTF is 21.34 which shows that it outperforms other methods with ARIMA-NN being the next best at 25.53 MAPE (%). This shows that IMLEA-STLF shows an improvement of 16.4% compared to the next best approach. Electricity price forecasting is crucial for market participants to develop advanced bidding strategies, particularly in understanding competitors' bidding curves, captured by the residual demand curve (RDC). This curve expresses the market-clearing price as a function of the quantity of energy a company is willing to buy or sell. Traditional approaches to RDC estimation have limitations in forecasting accuracy and computational efficiency.

Recent literature includes novel hybrid approaches for forecasting such as using deep learning techniques, specifically long short-term memory (LSTM) networks, to forecast day-ahead RDCs by transforming temporal RDC sequences into a sequence of images, thus avoiding feature reduction [22]. A hybrid LSTM approach in combination with one-dimensional and convolutional LSTM networks incorporates exogenous information, such as renewable energy forecasts, into RDC forecasting. By leveraging GPUs and online learning, the computational time is reduced, making this possible for industrial applications. The results demonstrate the effectiveness of the hybrid LSTM approach, achieving excellent day-ahead forecasts with significant improvements in both root mean square error (RMSE) and Frèchet distance compared to benchmark models.

Future research directions include enhancing interpretability for decision-makers by analyzing the effect of exogenous variables on curve shape and modeling the impact of complex bidding conditions on clearing prices and curves.

3.3. Load Forecasting

In the rapidly evolving landscape of electricity markets, accurate short-term load forecasting (STLF), probabilistic load forecasting and long-term hourly load forecasting are critical for operational efficiency, resource management, and demand-side response facilitation. Various advanced machine learning and deep learning techniques have emerged to address the limitations of traditional methods, which often struggle with nonlinear relationships and complex patterns in load data.

Probabilistic load forecasting has also gained prominence, addressing the inherent uncertainties in electricity demand that traditional point forecasting methods often overlook. Techniques like the hybrid method combining Generalized Extreme Learning Machine (GELM) and Improved Wavelet Neural Networks (IWNN) have shown promise in providing reliable probabilistic intervals by accounting for data noise and model uncertainties. Lastly, long-term hourly load forecasting, essential for generation-transmission expansion planning, benefits from integrating machine learning regression methods with fuzzy logic-based inference engines.

Similarly, Generative and Moving Interactive Neural Networks (GMINN) [23] are also use for accurate short-term load forecasting in electricity markets. GMINN framework addresses several challenges faced by existing models, such as overfitting, parameter sensitivity, modeling complexity, and data leakage. GMINN integrates multiple components, including Mixup for data augmentation, moving average filter for decomposition, sample convolution with interactive learning, and genetic algorithm for hyperparameter optimization. Mixup expands the training dataset by generating new samples, while the moving average filter decomposes feature sequences into trend and residual sequences, preventing data leakage. Sample convolution with interactive learning captures time dependencies and models the sequences for forecasting. The genetic algorithm optimizes hyper-parameters, enhancing the framework's performance. Their experimental results on four real datasets from US electricity markets demonstrating that GMINN outperforms other baseline models in terms of mean absolute percentage error (MAPE) shown in Table 3 Compared with the second-best model, GMINN improves the MAPE by 4.0%.

This shows its effectiveness in short-term load forecasting. Ablation experiments validate the importance of each component in GMINN, providing insights for future research directions. Despite the positive results, GMINN requires significant computational resources, and the hyper-parameter search process is time-consuming, which may limit its practical applicability. Future work will aim to explore more efficient model architectures and optimization techniques to mitigate these challenges and apply the framework to additional real-world scenarios.

Multi-Scale Ensemble Method (MSEM) [24] is another approach to enhance the accuracy of short-term load forecasting (STLF) in electricity markets. This method leverages the power of ensemble learning, combining Stochastic Weight Averaging (SWA) and Differential Evolution (DE) ensemble techniques to improve the generalization ability of deep learning models. In Multi-Scale Ensemble Neural Network (MSENN), deep learning models such as Long Short-Term Memory (LSTM), Gate Recurrent Units (GRU), and Temporal Convolutional Networks (TCN) are integrated as basic models. The MSEM is applied at both single-model scale and multi-model scale, effectively reducing the generalization error and achieving high-precision fitting of future load samples. Their experimental results on load data from Hubei Province, China, show that the proposed MSENN outperforms eleven other intelligent STLF models, including machine learning, deep learning, and ensemble deep learning models. Notably, the MSENN-D model, with the proposed approach exhibits the best overall prediction performance as shown in Figure 6. R square is used as a measure for model performances which is described in equation (4). An R-squared value shows how well the model predicts the outcome of the dependent variable.

$$R^2 = 1 - \frac{\Sigma_i (y_i - \hat{y}_i)^2}{\Sigma_i \left(y_i - \underline{y}\right)^2} \qquad (4)$$

Traditional methods based on statistics and machine learning often fail to capture the nonlinear relationships inherent in load data, prompting the adoption of deep learning techniques. However, Short-term load forecasting (STLF) is another method the context of an increasingly open electricity market, which focuses on accurate predictions for safe and stable grid operation, resource efficiency, and demand-side response facilitation. Qin Shen et al. [24] highlighted several key findings, the positive impact of meteorological factors on STLF accuracy, the effectiveness of model ensembling at any scale, and the complementary nature of simultaneous ensembling from multiple scales. However, it acknowledges the influence of additional factors such as human behavior on STLF accuracy, suggesting avenues for future research. Further research could explore the application of this technique to other domains such as power generation from photovoltaic and wind sources, considering additional influencing factors for enhanced forecasting accuracy. To address the limitations of existing methods, Yi Xuan et al., [25] proposed a two-step approach. Firstly, a feature selection algorithm based on random forest is employed to identify relevant input features for the load forecasting model. Subsequently, a hybrid neural network combining a Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (BiGRU) is developed. Multiple sub-models are trained using different width sliding time windows, and the final load forecast is obtained by averaging the predictions of these sub-models. The experimental results using load datasets from regions in New Zealand and Zhejiang, China [25], showing the superior accuracy of the proposed method compared to other forecasting models. Specifically, the proposed model outperforms single-model CNN-BiGRU predictions as well as other deep learning models in terms of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Recent studies also include hybrid methods for probabilistic electricity load forecasting [45] to addresses the challenges posed by the restructuring of electricity markets and the increasing demand for accurate and reliable load predictions such as the combination of Generalized Extreme Learning Machine (GELM), Improved Wavelet Neural Networks (IWNN), wavelet preprocessing, and bootstrapping to account for model uncertainties and data noise. This ultimately provides load probabilistic intervals. This method leverages the benefits of GELM and IWNN for improved load forecasting. GELM is employed to train IWNNs, offering advantages in terms of speed and accuracy compared to traditional neural networks. Additionally, wavelet preprocessing is utilized to divide load data into subsets. This allows for improved forecasting models tailored to the characteristics of each subset. Bootstrapping is then applied to estimate uncertainties and generate PIs, considering both model uncertainties and data noise.

Lastly, the experimental validation of the method using data from the Ontario and Australian electricity markets demonstrates its high performance, accuracy, and reliability. The inclusion of Friedman and post-hoc tests further validates the approach, highlighting its practical applicability and effectiveness in addressing the challenges of probabilistic load forecasting.

Recent literature suggests there is a crucial need for reliable long-term hourly load forecasting in the electricity sector, particularly for informing decision-making processes related to generation-transmission expansion planning [27]. To tackle this challenge, a data-driven knowledge-based system is developed, integrating advancements in both the knowledge base and inference engine. In the knowledge base, they incorporate two categories of driving factors. First is long-term trend-related features like macro-economic factors and the other is short-term factors such as temperature. These factors are decomposed into three steps to address their frequency differences effectively. Machine learning regression methods including support vector machines, random forests, and artificial neural networks are compared to capture the nonlinear relationships of variables in each step. The results from these steps are then combined to form the forecasting model. In the inference engine, a novel reasoning method based on fuzzy logic is introduced to forecast regional long-term hourly load under uncertain temperature conditions, a significant factor in long-term forecasting. Figure 7 shows the turning point MAPE for the proposed approach. The forecast system [40] is evaluated against five other systems using datasets from the ISO New England electricity market and Iran's residential, commercial, and agricultural electricity load. The results demonstrate the superiority of the developed system in terms of accuracy and efficiency.

Table 3. RMSE, MAE and MAPE of different models for load forecasting (Choujun Zhan et al. [23])

Model	RMSE	MAE	MAPE (%)
GRU	1159.497	970.744	5.904
LSTM	1105.416	891.899	5.468
ANN	1847.298	1553.936	10.058
Adaboost	1410.298	1176.420	7.593
KNN	1911.969	1600.316	9.747
LightGBM	1422.353	1196.008	7.667
Random Forest	1543.659	1313.911	8.13
XGBoost	3557.056	3186.340	21.497
SCINet	938.717	760.922	4.623
TCN	886.245	719.134	4.716
NBeats	952.976	756.811	4.639
BLS	3284.482	2452.275	15.550
GMINN	847.170	713.171	4.440

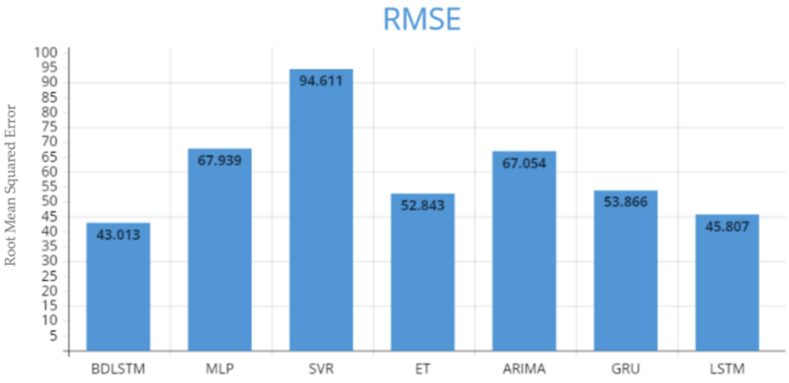

Figure 4. RMSE for BDLSTM and other approaches for price forecasting

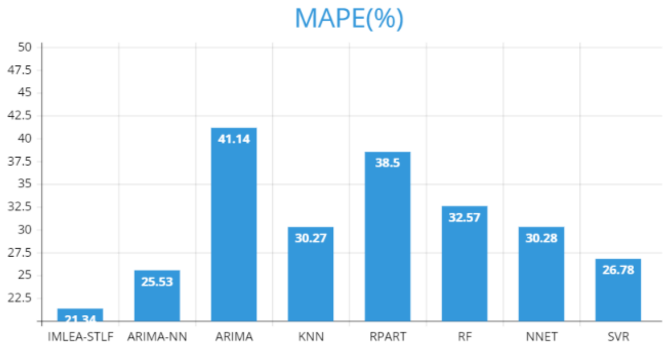

Figure 5. MAPE (%) of different approaches for demand forecasting

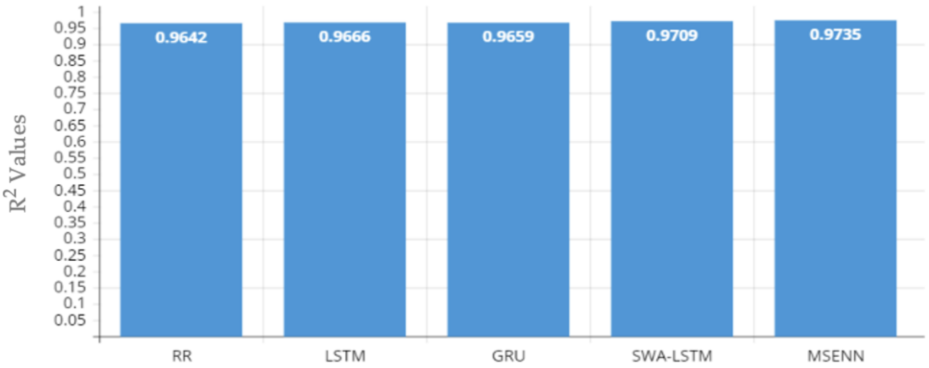

Figure 6. R-square values for the proposed model and other approaches for load forecasting

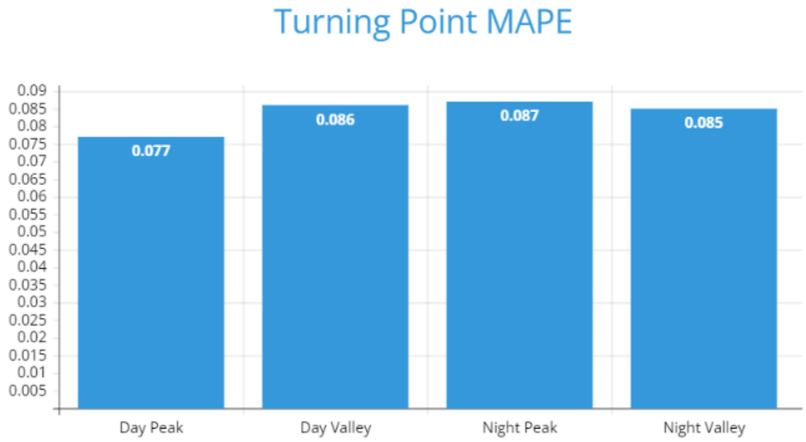

Figure 7. Turning Point MAPE for load forecasting

3.4. Carbon Emission Forecasting

The trends and current data have a crucial role in predicting the future emission of carbon dioxide and other greenhouse gasses. Carbon emission forecasting is crucial for regulatory compliance, environmental impact, economic efficiency, grid management, informed decision-making, consumer awareness, innovation and technology development. For effective emission forecasting, Particle Swarm Optimized-extremely randomized trees (PSO-ERT) regressor model [10] has been used for intensity calculation in the Australian National Electricity Market (NEM). Utilizing historical weather, demand, and generation data from the NEM regions, the PSO-ERT model outperforms other similar models such as Long Short Term Memory (LSTM) and Extreme Learning Machine (ELM), exhibiting Root Mean Squared Error (RMSE) values of 42.9, 109.8, 165.7, 25.4, and 74.1 for New South Wales, Tasmania, South Australia, Queensland, and Victoria, respectively.

Previous studies have primarily focused on long-term forecasts or employed complex neural network models like LSTM. However, simpler models such as PSO-ERT [10] address the complexity of emissions forecasting in the NEM properly, focusing on the role of different parameters in forecast accuracy utilizing machine learning techniques. Regions with higher renewable penetration exhibit different emission intensity patterns, highlighting the need for tailored approaches to forecasting. Additionally, it identifies weekends and times of lower operational demand as periods of reduced forecast accuracy, suggesting the potential benefits of incorporating pre-dispatch solar generation data into predictor feature sets. As the first study of its kind in the Australian NEM, Vahid Aryai et al. [10] set a precedent for future investigations in emissions intensity forecasting and contributed to the ongoing efforts towards emissions reduction and sustainable energy management.

3.5. Supply Forecasting

It is important to predict the amount of electricity that will be available from generation through different sources over a certain period. Traditional methods assume structural constraints such as boundedness and smoothness, which do not accurately represent the characteristics of real-world supply curves. Recognizing the importance of accurately analyzing these curves for both market participants and regulators, a hierarchical clustering method exist such as the work present by

Zehang Li et al. [11] to predict supply curves in electricity markets, addressing the limitations of existing techniques.

Key to the clustering-based supply forecasting is the incorporation of information on the price distribution of offers, enabling more accurate clustering and prediction of supply curves. By emphasizing differences in the most frequent price intervals, the weighted distance captures the characteristic jumps in the curves. The clustering results reveal insights into the temporal and market variables that influence the formation of supply curve clusters, providing valuable information for market analysis and decision-making.

Moreover, supervised classification procedure helps in identifying the most important variables for characterizing the clusters, facilitating the development of predictive models. The prediction process demonstrates the competitive performance of the proposed procedure across different markets, highlighting its potential for practical application.

4. Impact of renewable energy on price in the Smart Electricity Market

The integration of renewable energy resources into electricity markets has a profound impact on electricity prices. Data analytics can be used to understand and manage these effects. The variability in renewable energy resources can cause fluctuations in electricity prices. These can be analyzed using data analytic techniques which offer insights into how renewable energy impacts market prices.

It is important to understand the multi-directional impact of renewable energy on electricity markets, considering factors such as demand, operable power facility capacity, fuel prices, and seasonal variations. Traditional approaches to analyzing electricity markets often rely on pre-determined assumptions and fail to capture the dynamic interactions between variables. Traditional models often assume static relationships between price and supply-demand variables, which overlook the stochastic nature of renewable generation and its influence on market volatility. Advanced approaches such as ML and XAI provide the ability to model nonlinear interactions, integrate high-frequency data, and capture short-term fluctuations caused by weather conditions, demand spikes, and variations in generation capacity. These methods not only improve forecast accuracy but also enhance interpretability, enabling market operators and policymakers to identify drivers of price changes in real time.

Renewable energy has a huge impact on electricity market, which is comprehensively demonstrated by Mizue Shimomura et al. [6] focusing on the Japanese context. As the global transition towards decarbonization accelerates, understanding the dynamics of electricity markets influenced by RE becomes increasingly crucial. Complex challenges in the context can be addressed by the introduction of variable renewable energy (VRE), which not only affects the average market price through the merit order effect (MOE) but also introduces volatility and fluctuations due to its inherent variability [38]. The combination of machine learning and explainable artificial intelligence (XAI) can help uncover the intricate relationships within the Japanese electricity market to reveal significant insights into the drivers of market price fluctuations. While the MOE leads to a decline in prices during periods of high solar generation, the impact varies depending on factors such as time of day, season, and demand [6]. Demand emerges as the most influential factor, with prices exhibiting nonlinear increases during peak demand periods. Additionally, the interaction between solar generation, demand, and operable power facility capacity significantly impacts market volatility and price surges, particularly during periods of low solar generation and high demand, such as summer evenings [39].

Several studies highlight the importance of integrating flexible market mechanisms, demand response strategies, and energy storage solutions to mitigate the adverse price volatility caused by variable renewable energy [40, 41]. Furthermore, how data-driven approaches can support policy-making by quantifying the effects of renewable penetration on price behavior, grid stability, and overall market efficiency [42].

5. Trading in the Smart Electricity Market

A vital component of smart electricity markets is electricity trading. Data analytics can significantly improve market efficiency and sustainability. The electricity markets can be operated more efficiently by optimizing pricing mechanisms and reducing energy waste through data-driven trading strategies. The integration of experimental economics and machine learning in modeling trading behavior in electricity markets presents a promising approach to addressing the complexities of modern grid operations and market dynamics. Furthermore, technologies like blockchain can enhance transparency and security in trading. Here trading behavioral modeling and decentralized trading framework are discussed.

Two primary directions dominate research in trading analytics: (i) data-driven behavioral modeling, which uses machine learning and experimental economics to simulate and optimize trading behavior; and (ii) decentralized trading frameworks, which employ blockchain and game theory to enable secure and efficient peer-to-peer transactions. Hybrid models, such as those combining Generative Adversarial Networks (GANs) with interpretability frameworks (e.g., local interpretable model-agnostic explanations (LIME)), improve transparency and reliability in trading decisions such as Hybrid Experimental Learning (HEL) [4]. HEL aims to overcome the limitations of traditional analytical models and black-box machine learning approaches by providing both accuracy and interpretability. This offers a data-driven approach that minimizes assumptions while providing interpretability. By leveraging historical and experiment-simulated data, HEL models trading behavior using a GAN, which learns from human participants' behavior in experiments. Moreover, HEL incorporates an interpretation approach LIME to explain the output mechanism of the generative model. This ensures that the model's decisions are transparent and understandable, facilitating its application in practical scenarios where human decision-makers require trust and comprehension of the model's workings. Table 4 presents the studies focusing on trading in smart electricity markets.

Similarly, blockchain-based frameworks, coupled with machine learning algorithms, support adaptive bidding strategies and enhance profitability in decentralized energy markets. The integration of Connected Electric Vehicles (CEVs) into smart city infrastructure presents an opportunity to revolutionize energy management and trading systems. With their potential to reduce road congestion, enhance road safety, and contribute to environmental sustainability, CEVs are positioned to play a significant role in the future of transportation and energy. As CEVs become more prevalent, there is growing interest in leveraging their capabilities as mobile power plants to participate in electricity markets. With the anticipated growth of CEV adoption worldwide, there is a pressing need for innovative solutions to facilitate electricity trading and integration into the grid. Traditional centralized electricity systems pose limitations in terms of efficiency and privacy, necessitating decentralized approaches like the Decentralized Electricity Trading Framework (DETF) [46]. The use of blockchain technology, particularly consortium blockchain, offers a decentralized platform for CEVs to engage in peer-to-peer electricity trading without the need for intermediaries. This enhances trust, transparency, and security in transactions, addressing key concerns related to privacy and data exchange.

Furthermore, the incorporation of ML and game theoretic models enhances the efficiency and profitability of energy trading by enabling adaptive bidding strategies based on market dynamics and individual preferences. The proposed HLProfitX algorithm facilitates intelligent decision-making for CEVs, optimizing their energy transactions to maximize profitability. Numerical simulations conducted using MATLAB and Solidity demonstrate the effectiveness of the DETF and the HLProfitX algorithm in improving CEVs' profitability and energy trading management. Comparison with existing methods, such as PETCON, highlights the superiority of the proposed framework in terms of profitability and efficiency.

These approaches illustrate how advanced data analytics and emerging technologies can transform electricity trading by improving market efficiency, ensuring secure transactions, and providing interpretable insights for decision-makers. Future research is expected to expand these frameworks to incorporate renewable energy integration, real-time market conditions, and scalable blockchain solutions for large-scale deployment.

6. Bidding in the Smart Electricity Market

Bidding strategies in the smart electricity markets can be optimized by utilizing data analytics for better market outcomes and renewable energy integration. We can further demonstrate how these strategies improve market dynamics by analyzing data-driven bidding approaches. Virtual bidding, a financial position not backed by physical assets, plays a crucial role in enhancing market efficiency, reducing price spreads between day-ahead (DA) and real-time (RT) locational marginal prices (LMPs), and increasing market liquidity. However, existing approaches often underestimate profitability and overestimate market efficiencies by neglecting transaction fees and uplift costs, and failing to model the impacts of virtual bidders' trading activities on electricity market prices explicitly.

To address these limitations, Y. Li et al. [30] proposed an algorithmic virtual bid trading strategy that incorporates machine learning techniques to forecast LMP spreads and virtual trading quantities, and a constrained gradient boosting tree to model the sensitivity of LMP spread concerning net virtual bids. The virtual bid portfolio optimization framework, considering both risk constraints and price sensitivities, is developed to maximize profitability while accounting for market dynamics.

Empirical analysis conducted across multiple U.S. wholesale electricity markets (PJM, ISO-NE, and CAISO) demonstrates the effectiveness of the proposed approach. Results show that the algorithmic virtual bid trading strategy outperforms existing methods, achieving significant profits and higher Sharpe ratios for virtual bid portfolios. Particularly, the profitability is highest in CAISO, indicating potential opportunities for algorithmic trading strategies to exploit congestion pattern differences between DA and RT markets.

Dong Han et al. [31] addressed the challenges posed by the discrepancies between day-ahead (DA) and real-time (RT) prices in the two-settlement electricity market by proposing a virtual bidding (VB) strategy to arbitrage the price differences and promote price convergence. The model considers optimal bidding for virtual bidders in spatiotemporal dimensions, taking into account budget constraints and utilizing both decrements and increment bids to maximize cumulative payoff while hedging risks. Formulated as a Markov Decision Process (MDP) problem, the model incorporates the conditional value-at-risk (CVaR) to quantify and hedge risks faced by virtual bidders. To solve the optimal bidding strategy problem, a deep reinforcement learning (DRL) algorithm is employed, enabling effective decision-making through continuous interaction with a simulated environment.

The proposed methodology is evaluated using PJM data from 2016 to 2018, comparing it with greedy algorithms and dynamic programming methods. Results demonstrate the effectiveness and superiority of the DRL algorithm, showing higher cumulative profits and Sharpe ratios. Moreover, the analysis of different budget scenarios highlights the trade-off between cumulative profits and risks, further validating the efficacy of the proposed approach.

Strategic bidding in electricity markets is another critical issue, where power producers' profits are determined by market prices rather than regulated rates. Specifically, in the context of uniform pricing markets, where complexities arise from the interactions among participants, the effectiveness of prediction-based bidding strategies is explored. While prediction-based approaches, leveraging machine learning algorithms, are generally effective, they may sometimes yield unsatisfactory results due to the intricacies of the market environment.

To address this challenge, Yueyong Yang et al. [32] proposed a selective learning scheme for strategic bidding, based on an ensemble technique. Multiple machine learning algorithms are employed to predict market prices and generate bidding recommendations. As the market clearing iteration progresses, the most suitable algorithms are selected to guide bidding strategies. Notably, the prediction method used within the selective learning scheme is modified to enhance accuracy in capturing the market's characteristics. Simulation studies are conducted to validate the effectiveness of the scheme, demonstrating its ability to lead to more rational bidding behaviors and higher profits. This scheme, which comprises multiple complementary bidding algorithms, allows for the automatic selection of the most advantageous algorithm based on real-time performance. Figure 8 shows the Profit in terms of (10^3) RMB for different machine learning models.

The strategic bidding in uniform pricing electricity spot markets underscores several key insights: bidding behavior can impact the profits of both individual producers and the entire system; incorporating internal bidding data enhances prediction accuracy; comprehensive algorithms can lead to more appropriate bidding strategies and increased profits; while the selective learning scheme can boost a producer's profit, it may also increase system costs; finally, pay-as-bid pricing strategies may result in lower profits due to constraints on market power.

Table 4. Trading in Smart Electricity Markets

Study	Approach	Key Contribution
Wenxuan Liu et al. [4]	Hybrid Experimental Learning (HEL) with GAN + LIME	Combines behavioral modeling with interpretability to optimize trading decisions.
Dhaou Said [29]	Decentralized Electricity Trading Framework (DETF) with Blockchain + ML + Game Theory	Enables secure peer-to-peer energy trading, improving profitability and grid stability.
Y. Li et al. [30]	ML-based Virtual Bid Trading Strategy with Gradient Boosting Trees	Models locational marginal price (LMP) spreads and virtual bid quantities to maximize trading profits.
Dong Han et al. [31]	Deep Reinforcement Learning (DRL) for Virtual Bidding	Uses DRL and CVaR for risk-aware bidding strategies, improving cumulative profits in two-settlement markets.
Yueyong Yang et al. [32]	Selective Learning with Ensemble ML Models	Selects best-performing prediction models dynamically to improve bidding recommendations and profitability.

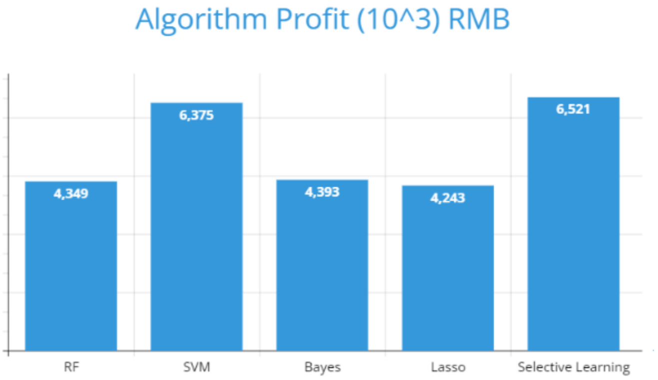

Figure 8. Algorithm Profits for Different Models for bidding recommendation in smart electricity markets

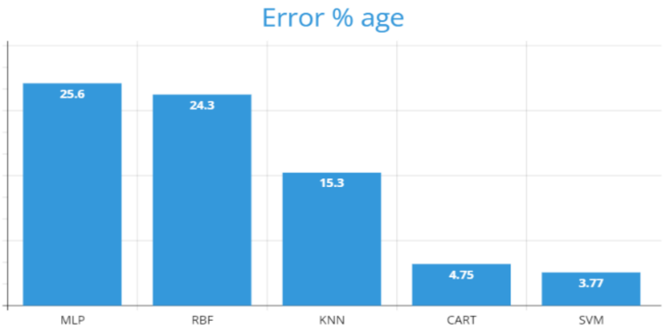

Figure 9. Error % age for collusion detection using different models

7. Security in the Smart Electricity Market

Security in smart electricity markets is important for maintaining the efficiency and reliability of systems that integrate renewable energy sources. Smart microgrids (SMGs) have great importance in areas where traditional transmission lines are impractical, such as remote islands or isolated communication stops. SMGs, incorporating renewable energy sources (RESs), storage systems, and electronic loads, function independently of the main power grid, enhancing efficiency and resilience [44]. However, the convergence of cyber and physical systems in SMGs exposes them to cyber threats, with potential consequences for system security and operation. Past assumptions about the invulnerability of industrial control systems to cyber-attacks have been challenged, necessitating robust cybersecurity measures for SMGs. These threats include false data injection and collusion. False data injection refers to a cyber-attack where erroneous data is intentionally inserted into a system to manipulate its operation and output like load forecast manipulation, price manipulation and system disruption. Yikun Huang et al. [33] identified the need for effective detection methods for False Data Injection Attacks (FDIAs) in AC-based SMGs, given the complexity of nonlinear systems and the potential for cyber-attacks to disrupt normal operations. In the context of increasing reliance on renewable energy sources (RESs) and the deployment of off-grid MGs, the security of SMGs becomes paramount, encompassing both physical and cyber layers. While existing approaches focus on linearization or specific attack scenarios, the proposed solution employs deep learning algorithms, specifically Deep Belief Networks (DBNs), to detect FDIAs in real-time. Table 5 shows the detection layout reaction (%) for the false data injection.

By utilizing historic measurement data and specifications for attack recognition, the proposed method demonstrates high precision in detecting FDIAs, even in the presence of load variations and cyber-attacks. By leveraging wavelet singular values and deep learning algorithms, this method achieved a detection precision of over 97.5%, outperforming conventional techniques.

Collusion is another type of cyber-attack where two or more parties conspire to manipulate a system for mutual benefit. Identifying such an attack refers to collusion detection.

Table 5. Detection Layout Reaction(%) for FDIAs Confusion Matrix (Yikun Huang et al. [33])

Real Amount	Positives	Negative
Positives	97.57	1.65
Negatives	2.43	98.35

Peyman Razmi et al. [34] presented an approach for detecting collusion in electricity markets using supervised machine learning techniques. In competitive electricity markets, producers aim to optimize their production levels based on market prices. However, in oligopoly markets, producers may engage in collusion to manipulate prices in their favor, resulting in market inefficiencies and consumer harm. Collusion can take various forms, including implicit and explicit agreements among generation firms.

Detecting collusion is challenging due to the lack of specific data directly referring to collusion behaviors. While supervised learning offers a potential solution, Peyman Razmi et al. [34] explored the use of machine learning algorithms such as classification and regression tree (CART) and support vector machine (SVM) to detect collusion based on day-ahead market data. By identifying possible collusion scenarios and computing market equilibriums under different scenarios, it trains the collusion detection machine using data from equilibrium points and their peripherals. Simulation results demonstrate the effectiveness of this approach in detecting collusion, with SVM and decision tree algorithms showing the highest accuracy among the tested methods. This is shown in Figure 9.

Furthermore, the study evaluates the impact of different criteria on collusion detection accuracy, highlighting the importance of certain indices such as market clearing (MC) and load error rate (LER) in classification performance. Independent system operators (ISOs) can run the collusion detection program periodically, and persistent collusion over time indicates the validity of collusion suspicions.

Technological advancement has given massive rise to cyber-attacks and the need to solve them. The common cyber-attacks in the smart electricity market are false data injection and collusion detection and thus, there is a need to detect these attacks to prevent any harm. The approaches proposed have provided security in the field of the smart electricity market. However, cyber-attacks are revolutionizing along with technological advancements and so these approaches need to be updated to detect any cyber-attack that might arise.

8. Future Perspectives

This systematic review comprehensively investigates data analytics techniques in smart electricity markets. This analysis is done for 23 selected research papers from a pool of 504 papers. The review identifies several key findings and trends that contribute to advancing the field. During the introduction part, research questions were defined that were aimed to be answered in the paper and a discussion on the research questions is covered in this section. Moreover, future gaps are identified, and advancements are discussed to counter these future gaps.

In this systematic review, it is clearly seen that there are many different data analytics techniques for analyzing smart electricity markets. For forecasting, we can see machine learning algorithms like XG-Boost, Support Vector Machines, and Random Forest. Moreover, approaches like LSTM, GRU and neural networks like IWNN and WNN. Hybrid models including CNN, BiGRU and BiLSTM are also utilized. These have a better ability to handle nonlinear and complex relationships in data. Impact of Renewable energy on price utilized machine learning and explainable AI (XAI). Furthermore, trading observed Hybrid Experimental Learning and hybrid models combining blockchain technology, machine learning and game thoracic models. For bidding, traditional machine learning models, selective learning and deep reinforcement learning played vital roles. Lastly, for security deep belief networks, Support Vector Machines and classification and regression tree (CART) were utilized. We can see the variety of novel data analytic techniques across different fields of application in smart electricity markets.

The literature that in future the integration of diverse data analytics approaches across forecasting, trading, bidding, and security is expected to drive the evolution of smart electricity markets toward autonomous, resilient, and fully data-driven ecosystems. Studies such as those by Yueyong Yang et al. [32] and Y. Li et al. [30] demonstrate how ensemble and gradient boosting models can optimize bidding strategies, while Dong Han et al. [31] highlights the potential of deep reinforcement learning for risk-aware trading. Similarly, the hybrid frameworks proposed by Wenxuan Liu et al. [4] and Dhaou Said [29] point toward future platforms where interpretability, decentralization, and real-time adaptability converge. In the domain of price impact analysis, Mizue Shimomura et al. [6] showcase how Explainable AI can uncover intricate market dynamics, suggesting broader applicability across forecasting and demand-side management. By synthesizing these advancements, future research should aim to develop holistic AI-driven market ecosystems that integrate forecasting accuracy, secure decentralized trading, adaptive bidding, and proactive cybersecurity, creating a foundation for next-generation energy markets.

Data analytics approaches have a very important role in smart electricity markets especially in forecasting. The application of forecasting includes price, load, supply, demand, and carbon emission. Temporal dependencies and seasonality in data are important to capture. Techniques like machine learning models, LSTM and its variants are widely utilized and are very effective in this regard. Moreover, with recent advancements in machine learning and deep learning, there is a need to improve accuracy by mitigating biases and uncertainties. Ensemble methods and hybrid models combining different algorithms improve the accuracy like a hybrid model combining CatBoost for feature selection and BDLSTM for price forecasting and a hybrid LSTM approach in combination with one-dimensional and convolutional LSTM networks. Lastly, identifying relevant predictors for more robust models is also important where feature selection techniques like Random Forest, Binary Genetic Algorithm and Gaussian Process Regression are utilized.

Technological advancements have revolutionized emerging trends and advancement along with the constant need to improve along with the technological advancements. There are many emerging trends in data analytics for smart electricity markets. This includes the integration of renewable energy forecasting with forecasting models. This involves incorporating weather data, seasons, holidays and other external factors into predictive models to enhance accuracy and reliability. Moreover, emerging approaches such as Generative Adversarial Networks (GANs), explainable AI and blockchain technology are becoming the core of emerging trends and advancements. These approaches are being widely used across many different fields of applications and areas including smart electricity markets. Moreover, the need for dynamic market conditions and sudden demand fluctuations is being addressed by advancements in real-time analytics and event-driven forecasting models.

Future research in data analytics for smart electricity markets should focus on developing real-time adaptive models capable of processing high-frequency data from decentralized energy systems. Integrating reinforcement learning and online learning algorithms, as explored by Dong Han et al. [31] for dynamic bidding, can enable fast decision-making under rapidly changing market conditions. Similarly, digital twin frameworks and generative models, such as those introduced by Wenxuan Liu et al. [4], may support predictive simulations of complex market behaviors. Probabilistic forecasting techniques, as demonstrated by C. Zhang and Y. Fu [18], should also be advanced to better quantify uncertainty, improve risk assessment, and enhance operational planning in volatile energy markets.

The recent advancements have caused the constant need for future advancements that keep on changing and becoming more complex. AI and Machine learning have become the core for future advancement for almost all fields of applications. Future advancements in smart electricity markets are poised to leverage advancements in AI and machine learning for more accurate and adaptive models. There is a lot of promise in techniques such as meta-learning for model selection and optimization, and reinforcement learning for adaptive decision-making in energy trading. Lastly, the integration of blockchain technology for decentralized trading and IoT for real-time data acquisition and processing is expected to reshape market operations and efficiency. Given the trend of advancements these potential future advancements will be constantly changing and the need for advancements will remain a crucial aspect and need to be revised constantly.

Another key direction is the integration of Explainable AI (XAI) to improve interpretability and trust in predictive models, particularly in high-stakes applications such as trading, bidding, and cybersecurity. The work of Mizue Shimomura et al. [6] highlights the value of XAI in understanding price fluctuations, providing a foundation for broader adoption in other domains of the smart electricity market. Moreover, scalable blockchain-based platforms for decentralized trading, such as the DETF proposed by Dhaou Said [29], should be extended to large-scale applications, incorporating peer-to-peer energy exchange and secure data management. Hybrid architectures combining AI, IoT, and edge computing may further enhance real-time monitoring, data-driven optimization, and predictive accuracy. Lastly, cybersecurity remains a critical area for future research, with promising directions including adversarial learning and deep belief networks, as shown by Yikun Huang et al. [33], to detect increasingly sophisticated attacks.

Furthermore, models can be enriched by incorporating additional data streams such as weather patterns, socio-economic factors, and grid infrastructure data. This holistic approach can capture complex interdependencies. This will also improve the robustness of predictions in dynamic market environments. It is essential to develop adaptive systems capable of adjusting to real-time data inputs and market conditions. This involves exploring online learning algorithms and reinforcement learning frameworks to optimize decision-making processes and enhance market responsiveness.

Moreover, it is also important to enhance transparency, security, and efficiency in energy transactions. Exploration of blockchain-based decentralized trading platforms can play a vital role in this regard. Future research should focus on scaling these solutions. These solutions should be integrated with existing market frameworks for broader adoption and impact. By addressing these avenues for future research, the smart electricity market can evolve towards more resilient, efficient, and sustainable energy systems, driven by advanced data analytics techniques and technological innovations.

5. Conclusion

This study provides a comprehensive review of data analytics applications in smart electricity markets, highlighting how advanced analytical techniques address the increasing complexity of modern energy systems. The findings reveal a clear shift from traditional statistical models to advanced machine learning, deep learning, and hybrid approaches, which significantly enhance forecasting accuracy across price, load, demand, supply, and carbon emission forecasting. Novel methodologies such as Explainable AI (XAI), blockchain-enabled decentralized trading frameworks, and hybrid experimental learning models are emerging as powerful tools to improve interpretability, transparency, and decision-making across various market functions.

The review also demonstrates that the integration of renewable energy profoundly influences price dynamics, with studies showing how XAI and machine learning uncover complex, non-linear relationships between generation variability, demand, and market volatility. In trading and bidding, advanced AI-based frameworks – including GANs, DRL, and ensemble methods – are enabling more accurate behavioral modeling and adaptive bidding strategies, while blockchain technology ensures secure and transparent transactions. Additionally, cybersecurity remains a critical concern, with studies presenting innovative detection mechanisms for false data injection and collusion attacks to safeguard market operations.

Overall, this study consolidates state-of-the-art research, identifies gaps in forecasting, trading, bidding, renewable integration, and security, and outlines potential avenues for future advancements. By synthesizing diverse approaches and highlighting novel technologies, the review underscores the pivotal role of data analytics in shaping the next generation of smart electricity markets. Future research focusing on real-time adaptive models, scalable decentralized platforms, and explainable AI solutions will further drive innovation, enhance market resilience, and accelerate the transition toward a sustainable and data-driven energy ecosystem.

Funding: “This research received no external funding”

Authors Contribution
Conceptualization, Muhammad M. Asim and Niusha Shafiabady; methodology, Muhammad M. Asim, Fareed Ud Din, Asif Karim; software, Muhammad M. Asim.; validation, Fareed Ud Din and Asif Karim; formal analysis, Fareed Ud Din, Niusha Shafiabady; investigation, Muhammad M. Asim and Fareed Ud Din.; resources, Niusha Shafiabady; data curation, Muhammad M. Asim; writing—original draft preparation, Muhammad M. Asim and Fareed Ud Din.; writing—review and editing, Asif Karim and Niusha Shafiabady; visualization, Muhammad M. Asim and Niusha Shafiabady; project administration, Fareed Ud Din. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials
All the reported data and cited in the paper and is available on request from the corresponding author.

Conflict of interests
The authors declare no conflicts of interest.

References

1. F. vom Scheidt, H. Medinová, N. Ludwig, B. Richter, P. Staudt, and C. Weinhardt, "Data analytics in the electricity sector – A quantitative and qualitative literature review," *Energy and AI*, vol. 1, p. 100009, Aug. 2020. DOI: https://doi.org/10.1016/j.egyai.2020.100009
2. Z. Qu, L. He, X. Ge, F. Wang, F. Xu, and J. Lu, "A Two-Stage Forecasting Approach for Day-Ahead Electricity Price Based on Improved Wavelet Neural Network With ELM Initialization," *IEEE Transactions on Industry Applications*, 2024. DOI: https://doi.org/10.1109/TIA.2024.3365456
3. T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustainable Cities and Society*, vol. 55, pp. 102052, Apr. 2020. DOI: https://doi.org/10.1016/j.scs.2020.102052
4. W. Liu, J. Zhao, J. Qiu, and Z. Y. Dong, "Interpretable Hybrid Experimental Learning for Trading Behavior Modeling in Electricity MarketTradib," *IEEE Transactions on Power Systems*, vol. 38, no. 2, pp. 1022-1032, 2023. DOI: https://doi.org/10.1109/TPWRS.2022.3173654
5. H. O. Riddervold, S. Riemer-Sørensen, P. Szederjesi, and M. Korpås, "A supervised learning approach for optimal selection of bidding strategies in reservoir hydro," *Electric Power Systems Research*, vol. 187, p. 106496, 2020. DOI: https://doi.org/10.1016/j.epsr.2020.106496
6. M. Shimomura, A. R. Keeley, K. Matsumoto, K. Tanaka, and S. Managi, "Beyond the merit order effect: Impact of the rapid expansion of renewable energy on electricity market price," *Renewable and Sustainable Energy Reviews*, vol. 189, p. 114037, 2024. DOI: https://doi.org/10.1016/j.rser.2023.114037

7. Y. Huang and H. He, "Advance learning technique for the electricity market attack detection," *Computers and Electrical Engineering*, vol. 100, p. 107865, 2022. DOI: https://doi.org/10.1016/j.compeleceng.2022.107865

8. Z. Qu, L. He, X. Ge, F. Wang, F. Xu, and J. Lu, "A Two-Stage Forecasting Approach for Day-Ahead Electricity Price Based on Improved Wavelet Neural Network With ELM Initialization," *IEEE Transactions on Industry Applications*, 2024. DOI: https://doi.org/10.1109/TIA.2024.3365456

9. X. Fang, J. Cui, T. Oozeki, and Y. Ueda, "Machine Learning-Based Japanese Spot Market Price Forecasting Considering the Solar Contribution," *IEEE Access*, vol. 12, pp. 52452-52465, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3387071

10. V. Aryai and M. Goldsworthy, "Day ahead carbon emission forecasting of the regional National Electricity Market using machine learning methods," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106314, 2023. DOI: https://doi.org/10.1016/j.engappai.2023.106314

11. Z. Li, A. M. Alonso, A. Elías, and J. M. Morales, "Clustering and forecasting of day-ahead electricity supply curves using a market-based distance," *International Journal of Electrical Power and Energy Systems*, vol. 158, p. 109977, 2024. DOI: https://doi.org/10.1016/j.ijepes.2024.109977

12. I. M. Mehedi, H. Bassi, M. J. Rawa, M. Ajour, A. Abusorrah, M. T. Vellingiri, Z. Salam, and M. P. B. Abdullah, "Intelligent machine learning with evolutionary algorithm based short term load forecasting in power systems," *IEEE Access*, vol. 9, pp. 100113-100124, 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3096918

13. P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region," *Energy*, vol. 185, pp. 1143-1153, 2019. DOI: https://doi.org/10.1016/j.energy.2019.07.074

14. Masood, Bilal, Fareed Ud Din, and Sobia Baig. "Role of narrow-band low data rate power line communication in smart grids and noise reduction strategy." World Applied Sciences Journal 26.12 (2013): 1595-1601.

15. Qadir, Z., Khan, Y. A., Rana, M. T. A., Din, F. U., & Shafiabady, N. (2024). Efficient energy utilization in smart grids: An artificial intelligence perspective. In Ethical Artificial Intelligence in Power Electronics (pp. 133-147). CRC Press.

16. A. Kheirandish, F. Motlagh, N. Shafiabady, M. Dahari, and A. K. A. Wahab, "Dynamic fuzzy cognitive network approach for modelling and control of PEM fuel cell for power electric bicycle system," *Appl. Energy*, vol. 207, pp. 591-603, 2017. DOI: https://doi.org/10.1016/j.apenergy.2017.05.084

17. Carrera-Rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). How-to conduct a systematic literature review: A quick guide for computer science research. MethodsX, 9, 101895. https://doi.org/10.1016/j.mex.2022.101895

18. C. Zhang and Y. Fu, "Probabilistic Electricity Price Forecast With Optimal Prediction Interval," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 442-452, Dec. 2023. DOI: https://doi.org/10.1109/TPWRS.2023.3235193

19. C. Xiao, D. Sutanto, K. M. Muttaqi, M. Zhang, K. Meng, and Z. Y. Dong, "Online Sequential Extreme Learning Machine Algorithm for Better Predispatch Electricity Price Forecasting Grids," *IEEE Transactions on Industry Applications*, vol. 57, no. 2, pp. 1860-1871, 2021. DOI: https://doi.org/10.1109/TIA.2021.3051105

20. W. Sai, Z. Pan, S. Liu, Z. Jiao, Z. Zhong, B. Miao, and S. H. Chan, "Event-driven forecasting of wholesale electricity price and frequency regulation price using machine learning algorithms," *Applied Energy*, vol. 352, p. 121989, 2023. DOI: https://doi.org/10.1016/j.apenergy.2023.121989

21. A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, and R. John Millar, "Machine Learning Based Integrated Feature Selection Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems," *IEEE Access*, vol. 7, pp. 91463-91475, Jul. 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2924685

22. A. Coronati, J. R. Andrade, and R. J. Bessa, "A deep learning method for forecasting residual market curves," *Electric Power Systems Research*, vol. 190, p. 106756, 2021. DOI: https://doi.org/10.1016/j.epsr.2020.106756

23. C. Zhan, D. Yin, Y. Shen, and T. Hao, "GMINN: A Generative Moving Interactive Neural Network for Enhanced Short-Term Load Forecasting in Modern Electricity Markets," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 1-1, 2024. DOI: https://doi.org/10.1109/TCE.2024.3367885

24. Q. Shen, L. Mo, G. Liu, J. Zhou, Y. Zhang, and P. Ren, "Short-Term Load Forecasting Based on Multi-Scale Ensemble Deep Learning Neural Network," *IEEE Access*, vol. 11, pp. 111963-111975, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3322167

25. Y. Xuan, W. Si, J. Zhu, Z. Sun, J. Zhao, M. Xu, and S. Xu, "Multi-Model Fusion Short-Term Load Forecasting Based on Random Forest Feature Selection and Hybrid Neural Network," *IEEE Access*, vol. 9, pp. 9321361-69009, 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3051337

26. M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-Khah, and J. P. S. Catalão, "Probabilistic Load Forecasting Using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6961-6971, Oct. 2018. DOI: https://doi.org/10.1109/TSG.2018.2807845

27. M. R. N. Kalhori, I. T. Emami, F. Fallahi, and M. Tabarzadi, "A data-driven knowledge-based system with reasoning under uncertain evidence for regional long-term hourly load forecasting," *Applied Energy*, vol. 314, p. 118975, 2022. DOI: https://doi.org/10.1016/j.apenergy.2022.118975

28. Li, Z., Alonso, A. M., Elías, A., & Morales, J. M. (2024). Clustering and forecasting of day-ahead electricity supply curves using a market-based distance. International Journal of Electrical Power & Energy Systems, 158, 109977.

29. D. Said, "A Decentralized Electricity Trading Framework (DETF) for Connected EVs: A Block-chain and Machine Learning for Profit Margin Optimization," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6594-6602, 2021. DOI: https://doi.org/10.1109/TII.2020.3045011

30. Y. Li, N. Yu, and W. Wang, "Machine Learning-Driven Virtual Bidding with Electricity Market Efficiency Analysis," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 354-364, 2022. DOI: https://doi.org/10.1109/TPWRS.2021.3096469

31. D. Han, W. Huang, H. Ren, W. Zhao, and Y. Li, "Machine learning analytics for virtual bidding in the electricity market," *International Journal of Electrical Power and Energy Systems*, vol. 143, p. 108489, 2022.
DOI: https://doi.org/10.1016/j.ijepes.2022.108489

32. Y. Yang, T. Ji, and Z. Jing, "Selective learning for strategic bidding in uniform pricing electricity spot market," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 6, pp. 1334-1344, 2021.
DOI: https://doi.org/10.17775/CSEEJPES.2020.02600

33. Y. Huang and H. He, "Advance learning technique for the electricity market attack detection," *Computers and Electrical Engineering*, vol. 100, p. 107865, 2022.
DOI: https://doi.org/10.1016/j.compeleceng.2022.107865

34. P. Razmi, M. O. Buygi, and M. Esmalifalak, "Collusion Strategy Investigation and Detection for Generation Units in Electricity Market Using Supervised Learning Paradigm," *IEEE Systems Journal*, vol. 15, no. 1, pp. 146-157, 2021.
DOI: https://doi.org/10.1109/JSYST.2020.2991608

35. M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Advanced Engineering Informatics*, vol. 35, pp. 1-16, 2018.
DOI: https://doi.org/10.1016/j.aei.2017.11.002

36. F. Zhang, H. Fleyeh, and C. Bales, "A hybrid model based on bidirectional long short-term memory neural network and Catboost for short-term electricity spot price forecasting," *Journal of the Operational Research Society*, vol. 73, no. 2, pp. 301-325, 2022.

37. C. Fraunholz, E. Kraft, D. Keles, and W. Fichtner, "Advanced price forecasting in agent-based electricity market simulation," *Applied Energy*, vol. 290, p. 116688, 2021.
DOI: https://doi.org/10.1016/j.apenergy.2021.116688

38. R. Subramanya, S. Sierla, and V. Vyatkin, "From DevOps to MLOps: Overview and Application to Electricity Market Forecasting," *Applied Sciences (Switzerland)*, vol. 12, no. 19, pp. 9851, 2022.
DOI: https://doi.org/10.3390/app12199851

39. N. Nizharadze, A. Farokhi Soofi, and S. Manshadi, "Predicting the Gap in the Day-Ahead and Real-Time Market Prices Leveraging Exogenous Weather Data," *Algorithms*, vol. 16, no. 11, pp. 508, 2023.
DOI: https://doi.org/10.3390/a16110508

40. M. A. Munoz, J. M. Morales, and S. Pineda, "Feature-Driven Improvement of Renewable Energy Forecasting and Trading," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3753-3763, 2020.
DOI: https://doi.org/10.1109/TPWRS.2020.2975246

41. T. Carriere and G. Kariniotakis, "An Integrated Approach for Value-Oriented Energy Forecasting and Data-Driven Decision-Making Application to Renewable Energy Trading," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6933-6944, 2019.
DOI: https://doi.org/10.1109/TSG.2019.2914379

42. J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, vol. 81, pt. 1, pp. 1548–1568, Jan. 2018.
DOI: https://doi.org/10.1016/j.rser.2017.05.234

43. J. M. Luna-Romera, M. Carranza-García, A. Arcos-Vargas, and J. C. Riquelme-Santos, "An empirical analysis of the relationship among price, demand and CO2 emissions in the Spanish electricity market," *Heliyon*, vol. 10, no. 3, p. e25838, 2024.
DOI: https://doi.org/10.1016/j.heliyon.2024.e25838

44. Masood, B., Usman, M., Din, F. U., & Haider, A. Effect of transient and non-transient models on the performance of PLC. Telecommunication Systems, 65(1), 55-64, 2017.

45. Laitsos, V., Vontzos, G., Paraschoudis, P., Tsampasis, E., Bargiotas, D., & Tsoukalas, L. H. (2024). The state of the art electricity load and price forecasting for the modern wholesale electricity market. Energies, 17(22), 5797.

46. Said, D. (2020). A decentralized electricity trading framework (DETF) for connected EVs: A blockchain and machine learning for profit margin optimization. IEEE Transactions on Industrial Informatics, 17(10), 6594-6602.