Renewable and Sustainable Energy Reviews 227 (2026) 116500

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Machine learning in peak demand forecasting: foundations, trends, and insights

Shuang Dai [a,*], Fanlin Meng [b], Hongsheng Dai [c], Qian Wang [d], Xizhong Chen [e], Wenlei Bai [f], Peizhi Shi [g], Richard Allmendinger [h], Yuchen Zhang [i,j], Jian Liu [k]

[a] Department of Engineering, University of Exeter, Exeter, UK
[b] University of Exeter Business School, University of Exeter, Exeter, UK
[c] School of Mathematics, Statistics and Physics, Newcastle University, Newcastle, UK
[d] Department of Computer Science, Durham University, Durham, UK
[e] School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai, China
[f] Electric Reliability Council of Texas (ERCOT), Taylor, Tx 76574, USA
[g] Centre for Decision Research, Leeds University Business School, University of Leeds, Leeds, UK
[h] Alliance Manchester Business School, University of Manchester, Manchester, UK
[i] Institute for Analytics and Data Science, University of Essex, Colchester, UK
[j] School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK
[k] Kummer Institute Center for Artificial Intelligence and Autonomous Systems, Missouri University of Science and Technology, Rolla, USA

HIGHLIGHTS

- First comprehensive review of peak demand forecasting, analyzing 186 studies.
- Categorized studies into three stages and defined a unified forecasting framework.
- Analyzed the evolving role of machine learning in different application contexts.
- Identified key challenges and outlined emerging trends for future research.

ARTICLE INFO

Keywords:
Deep learning
Machine learning
Peak demand forecasting
Power system
Smart grid
Time-series analysis

ABSTRACT

Peak demand forecasting involves predicting the maximum electricity demand within a specific period, which plays a key role in maintaining the efficiency and stability of power systems. The rapid evolution of power systems, driven by advanced metering infrastructure, local energy applications such as electric vehicles, and the increasing adoption of intermittent renewable energy, has introduced greater randomness and reduced predictability in peak demand. Given the pressing need to address more diverse implementation requirements across different contexts, accurate and reliable peak demand forecasting has become increasingly important. To the best of our knowledge, this study is the first to provide a comprehensive overview of peak demand forecasting methods. It systematically reviews 186 studies published since the 1950s, categorizing these methods into three stages based on their developmental timeline. Building on this, the study defines a unified framework for peak demand forecasting and offers an in-depth analysis linking these methods to the practical needs of power systems. Notably, it highlights the growing importance of machine learning-driven forecasting models in addressing the increasing complexity of modern energy environments. Furthermore, this study identifies key research gaps and points out emerging trends that hold potential for advancing innovation in this field.

* Corresponding author.
Email address: s.dai@exeter.ac.uk (S. Dai).

https://doi.org/10.1016/j.rser.2025.116500
Received 6 January 2025; Received in revised form 19 October 2025; Accepted 14 November 2025
Available online 26 November 2025

Nomenclature

Abbreviations

AdaBoost Adaptive Boosting
AI Artificial Intelligence
AMI Advanced Metering Infrastructure
ANN Artificial Neural Network
ARIMA Autoregressive Integrated Moving Average
AUC Computed Area Under the Curve
CPI Consumer Price Index
CLARA Clustering Large Applications
CNN Convolutional Neural Networks
cINN Conditional Invertible Neural Networks
CRPS Continuous Ranked Probability Score
DER Distributed Energy Resources
DSM Demand-Side Management
DTW Dynamic Time Warping
EVT Extreme Value Theory
GA Genetic Algorithms
GB Gradient Boosting
GNP Gross National Product
GRA Grey Relational Analysis
GRU Gated Recurrent Unit
GDP Gross Domestic Product
HVAC Heating, Ventilation, and Air Conditioning
HR Hit Rate
ICEEMDAN Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
IoT Internet of Things
KPI Key Performance Indicator
LSTM Long Short-Term Memory Networks
LTPDF Long-Term Peak Demand Forecast
MAE Mean Absolute Error
MAPE Mean Absolute Percentage Error
ML Machine Learning
MLP Multilayer Perceptrons
MSE Mean Squared Error
MTPDF Medium-Term Peak Demand Forecast
MI Mutual Information
PAPE Peak Absolute Percentage Error
PCA Principal Component Analysis
PICP Coverage Probability of Prediction Intervals
PV Photovoltaic
QRNN Quantile Regression Neural Networks
RBFN Radial Basis Function Networks
RFE Recursive Feature Elimination
RL Reinforcement Learning
RMSE Root Mean Squared Error
RNN Recurrent Neural Network
SARIMA Seasonal Autoregressive Integrated Moving Average
SHAP Shapley Additive Explanations
SOM Self-Organizing Maps
STPDF Short-Term Peak Demand Forecast
SVR Support Vector Regression
SVM Support Vector Machines
XGBoost Extreme Gradient Boosting

Symbols

$\hat{y}$ Predicted peak value
y Actual peak value
n Number of training samples
$\hat{t}$ Predicted peak time
t Actual peak time
ϵ Tolerance residual for peak time
h_i Binary flag indicating whether the predicted peak time falls within the tolerance interval

1. Introduction

The global energy landscape is undergoing a paradigm shift driven by the integration of renewable energy, Advanced Metering Infrastructure (AMI) [1], and the increasing reliance on data-centric solutions. Central to AMI systems are smart meters, which enable real-time data exchange between energy suppliers and end-users. As of mid-2024, the UK government reported that 36.2 million smart meters, comprising 63 % of all meters in Great Britain, had been installed [2]. Smart meters pave the way for smart energy applications such as Demand-Side Management (DSM). On the other hand, the integration of high-resolution energy consumption data with distributed intermittent energy resources such as wind, has introduced an unprecedented level of diversity and complexity to electricity demands [3].

Different electricity generation units have been adopted in power plants to meet these demands. Among all units, peak load power plants stand out for their economic inefficiency and environmental impact, as they operate exclusively during peak demand periods and are often highly polluting units. Sheffrin et al. [4] suggest that reducing peak loads by 5 %-15 % can produce substantial benefits, including resource optimization and lower real-time electricity tariffs. However, achieving these reductions requires effective strategies for managing peak demand. These strategies often leverage incentive and penalty mechanisms, including programs such as interruptible load control, demand-side bidding, and emergency demand response [5,6]. The central aspect of these strategies is accurately forecasting future peak demands, which guarantees optimal resource distribution and minimizes unnecessary costs and environmental impacts.

Peak demand forecasting differs from general load demand forecasting. While general load demand forecasting focuses on typical electricity consumption patterns driven by predictable daily or seasonal trends under stable system conditions, peak demand forecasting targets rare and extreme events that impose maximum stress on the power system. These events are highly sensitive to extreme factors, such as severe weather conditions [7], which can cause sudden and excessive spikes in electricity demand, necessitating specialized forecasting methods to maintain system reliability and manage resources effectively. Furthermore, evolving power systems, marked by the growing integration of distributed and intermittent energy sources such as wind and solar, along with advancements in energy storage technology, have brought new dynamics to peak load forecasting. These shifts have rendered traditional forecasting methods often inadequate, creating an urgent need for advanced techniques capable of addressing these dynamic challenges.

In response to these demands, Machine Learning (ML) has emerged as a transformative tool for peak demand forecasting in modern power systems. Modern power systems encompass diverse variables, including renewable energy generation patterns, energy storage dynamics, and real-time demand fluctuations. By integrating these variables, utilizing high-resolution data, and automating feature extraction, ML-driven methods offer dynamic and adaptive capabilities for efficient peak demand forecasting [8]. These ML-driven methods include the application of deep learning to capture complex relationships in peak demand, advanced probabilistic models to quantify uncertainty in forecasting, and hybrid approaches that combine ML techniques with traditional methods to improve precision and robustness [9]. Moreover, accurate peak

Table 1
The importance of peak demand forecasting for electricity stakeholders.

Electricity market stakeholders	Importance of peak demand forecast
Grid operators	• Improve the utilization rate of power generation equipment. • Reduce the cost of power generation and investment in power facilities. • Alleviate the supply pressure of the grid during peak hours [4].
Electricity retailers	• Make reasonable tariff schemes so as to maximize profits. • Offer energy-efficiency rebates to encourage customers to reduce peak demand [6].
Commercial and industrial end-users	• Improve the economic benefits and save production resources. • Alleviate environmental pollution by distributing emissions concentrated in the peak hours [10].
Residential end-users	• Save electricity bills and improve living standards [13].
Governments	• Enable a reliable power supply system. • Ensure economic growth and social welfare [11].

demand forecasting, achieved through these ML-driven techniques, is crucial for achieving both environmental and economic objectives. By anticipating demand spikes, system operators can activate demand-side measures, optimize dispatch from lower-emission sources, and avoid turning on carbon-intensive peaker plants that are often reserved for emergencies. Better foresight also enables more effective integration of intermittent renewables, reducing curtailment and balancing needs. For resource planners, accurate forecasts inform infrastructure investments by identifying true capacity requirements, preventing overbuilding and improving cost-efficiency. As summarized in Table 1, these forecasts address the specific needs and priorities of key stakeholders in the electricity market, including grid operators, electricity retailers, end-users, and governments, by providing tailored benefits that support their operational and strategic goals [10–12].

Given the increasing reliance on ML in power systems and its transformative impact on peak demand forecasting, a systematic review of its evolution and current state is essential. Despite growing interest in this area, no review has yet comprehensively addressed this specific domain. Recognizing this gap, this study examines the development of peak demand forecasting methods, focusing on the primary research question: *How have peak demand forecasting methods evolved from statistical approaches to ML-driven techniques, and how do these ML advancements address the challenges of modern power systems?* The review highlights methodological trends, identifies existing challenges, and explores pathways for future innovation in peak demand forecasting.

Unlike general load forecasting, which has been extensively reviewed in existing research and surveys [14–17], this review specifically narrows its scope to peak demand forecasting. Several related survey papers provide valuable insights into this area. For example, [3] discussed the role of various flexible resources in microgrids, highlighting their impact on operational strategies and the integration of renewable energy sources, both of which are essential for ensuring stable and reliable peak demand forecasting. [18] examined the role of demand response in managing forecast uncertainties, shedding light on the factors that influence peak demand forecasting. Furthermore, [19] reviewed artificial intelligence based strategies for load forecasting, highlighting the potential of hybrid and ensemble techniques for improving prediction accuracy in sustainable energy planning. By concentrating on peak demand forecasting, this study offers a thorough exploration that connects peak demand forecasting methods with the practical requirements of modern power systems. In the context of evolving power systems, the study further emphasizes the transformative role of ML-driven peak demand forecasting models in addressing the increasingly complex challenges encountered in dynamic and multifaceted environments.

The contributions of this study are fourfold. First, it provides a comprehensive analysis of the evolution of peak demand forecasting methods, emphasizing the transformative role of ML in addressing the complexities of modern power systems. Second, it establishes a unified framework to define peak demand forecasting, synthesized from the reviewed studies. This framework serves as a reference for future research and provides a foundation to address the specific challenges of this domain. Third, this study thoroughly analyzes the contexts of application, advantages and limitations of various ML-driven forecasting methods, providing valuable insights into their practical utility. Finally, it identifies under-explored opportunities for integrating emerging technologies and offers a roadmap for future research.

The remainder of this paper is organized as follows: Section 2 details the systematic review approach utilized and offers a summary of key findings from the literature, including a unified framework for peak demand forecasting. Section 3 highlights the challenges faced in early power systems and reviews traditional peak demand forecasting methods along with their limitations. Section 4 traces the evolution of peak demand forecasting methods and discusses significant breakthroughs in the field. Section 5 explores the contemporary trends in peak demand forecasting methods. This section also examines these methods across various application contexts and reflects on the advancements and current limitations of peak demand forecasting. Section 6 presents future research directions in this field, identifying current research gaps and pointing out emerging trends. Section 7 concludes this study by summarizing the key findings.

2. Research methodology and primary findings

The focus of this study is on the evolving role of ML in peak demand forecasting to address emerging challenges in power systems. Unlike general load forecasting, which focuses on forecasting electricity usage patterns under stable conditions, peak demand forecasting emphasizes accurately predicting extreme values. This involves accounting for the interplay of complex factors and addressing uncertainties inherent in predictions across different forecasting contexts. This study positions peak demand forecasting as both a distinct and integral component of general load forecasting, examining the development of the field and how ML has transformed it by providing more precise, scalable, and adaptable forecasting methods.

To achieve this, the review adopts a systematic strategy aimed at obtaining meaningful insights into the evolution of peak demand forecasting methods, their practical applications, and the factors influencing their effectiveness. This approach includes defining the scope of the study, selecting and analyzing relevant literature, synthesizing the findings into a unified framework, categorizing the selected research into three distinct stages based on their publication dates, conducting an in-depth analysis of each stage tracing the evolving role of ML in peak demand forecasting, and finally offering insights into future research directions.

2.1. Systematic review strategy

Table 2 details the methodology adopted to search and identify the relevant literature. Our search strategy was designed to systematically identify peer-reviewed studies on peak demand forecasting. For comprehensive and reliable coverage, the literature search was conducted across three major scholarly databases: Scopus, Web of Science, and Google Scholar. Scopus was chosen for its extensive coverage of engineering, energy, and applied science journals and conferences. Web of Science was included because of its coverage of core journals in energy systems and operations research. Google Scholar was used

Table 2
Strategy for identification and selection of relevant papers.

1 – Topic	The evolving role of ML in peak demand forecasting for power systems
2 – Research questions	**Primary research question:** How have peak demand forecasting methods evolved from statistical approaches to ML-driven techniques, and how do these ML advancements address the challenges of modern power systems? **Sub-research questions:** – *Fundamental context:* What methodologies dominated peak demand forecasting before ML, and how did they address early power system challenges? – *Transition to ML:* What breakthroughs have facilitated the shift to ML-driven peak demand forecasting? – *Contemporary trends:* What are the most effective ML techniques in peak demand forecasting, how do they vary across different contexts, and what are their key advantages and limitations? – *Future directions:* How can emerging trends in ML support the development of sustainable and adaptable energy solutions?
3 – Searching keywords	('peak' OR 'maximum') AND ('demand' OR 'load' OR 'load demand') AND ('forecasting' OR 'estimation' OR 'prediction')
4 – Searching database	Google Scholar, Scopus and the Web of Science
5 – Selection criteria	– Papers published in English in peer-reviewed journals and conferences were included. – Papers focused on methods specific to peak demand forecasting were included. – Papers focused on the comparison of different methods were included.

to complement these sources by offering the widest coverage across disciplines and publishers.

The search queries were tailored to focus on peak demand forecasting rather than general load forecasting. They were organized around three core concepts in the field: peak ('peak', 'maximum'), demand ('demand', 'load', 'load demand'), and forecasting ('forecasting', 'estimation', 'prediction'). These terms were combined using Boolean operators to form the query ('peak' OR 'maximum') AND ('demand' OR 'load' OR 'load demand') AND ('forecasting' OR 'estimation' OR 'prediction'). Broader terms such as 'load forecasting' or 'power demand forecasting' were deliberately excluded, as they tend to retrieve studies outside the scope of this review. Word-form variations were accounted for during the search and screening process, ensuring that differences in phrasing did not affect the results retrieved. For example, the query 'peak demand estimation' also retrieved results containing 'peak demand estimate' or 'estimating peak demand'. After records were retrieved, we employed a two-stage manual filtering process. First, we filtered out papers that were not written in English or not published in peer-reviewed venues. Second, we excluded papers that did not focus on methods specific to peak demand forecasting or on comparisons of different peak demand forecasting methods. In total, we obtained 186 papers published since 1950.

We must also recognize that our analysis might not reflect the latest developments in proprietary peak demand forecasting in the energy sector, such as confidential algorithms used for operational decision-making or trading. These advancements, which are not publicly accessible, probably comprise a significant share of cutting-edge knowledge in the domain. As a result, the findings presented here may not fully encompass the range of innovations in peak demand forecasting, especially in private and industrial environments.

2.2. Overview of findings

After applying the initial screening criteria, the core dataset for this review comprises 186 studies, all of which form the basis for future analyses.

Fig. 1 illustrates a word cloud summarizing the key methodologies employed in peak demand forecasting. Key terms such as 'ANN' (Artificial Neural Network), 'neural network,' and 'regression' dominate the word cloud, reflecting their central role in both traditional and contemporary forecasting techniques. Terms like 'LSTM' (Long Short-Term Memory Networks) and 'CNN' (Convolutional Neural Networks) highlight the increasing use of ML models in modern power systems. Meanwhile, the presence of 'ARIMA' (Autoregressive Integrated Moving

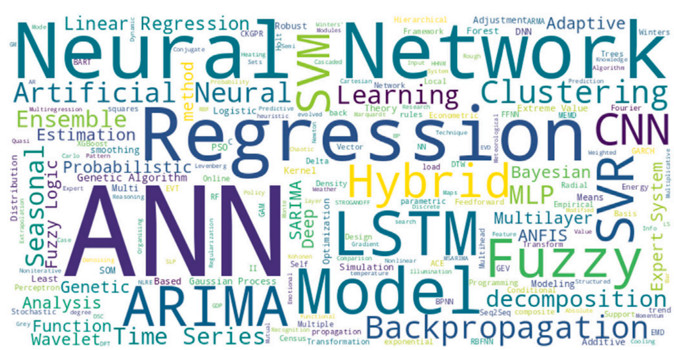

Fig. 1. Word cloud of peak demand forecasting methods based on 186 reviewed studies.

Average) highlights the continued significance of statistical time series models for analyzing historical data trends. Terms like 'hybrid' and 'fuzzy' emphasize the use of integrated and adaptive methods to enhance forecasting accuracy and robustness. Furthermore, terms such as 'clustering' and 'ensemble' reflect the application of specialized techniques to refine forecasting models.

The analysis of the reviewed studies shows a slow rise in publication activity up to 1990, a significant increase between 1991 and 2010, and steady growth from 2011 onward, highlighting the growing academic interest and continuous progress in the field. Based on these observations, the development of peak demand forecasting can be roughly segmented into three distinct stages:

- Pre-ML stage (up to 1990): Characterized by a limited research volume and reliance on traditional statistical methodologies such as linear regression and basic time series models.
- Transition to ML (1991-2010): Marked by the integration of hybrid models, improved computational tools, and the introduction of early ML techniques.
- ML stage (from 2011 onward): Defined by the broad use of ML and data-driven forecasting models that address the complexities of decentralized power systems and the growing role of renewable energy.

Note that the three stages are heuristic time bands defined by publication year. Because methods diffuse gradually, some papers near the boundaries display traits of adjacent stages. For example, statistical models remain widely used in the ML era, and early ML and hybrid

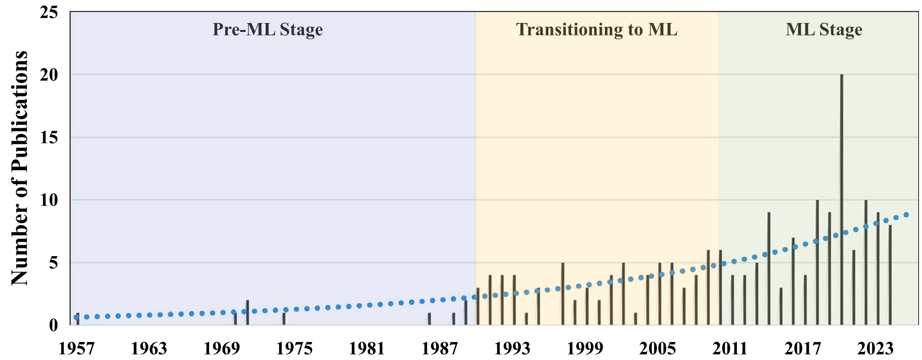

Fig. 2. Growth trend of publications and key stages in peak demand forecasting.

studies appeared during the 1990s. To manage overlap, we reference these studies in multiple sections but only tally each paper once at the stage level. Stage totals are therefore descriptive and should be considered along with the cross-stage hybridization patterns, which are further summarized in Table 8. To avoid double counting, we keep counts by publication year and focus our conclusions on directional trends such as the growth of ML use and increasing hybridization, which do not depend on strict boundary choices.

Fig. 2 shows the publication trend of research output across these three stages of peak demand forecasting. Specifically, from 1957 to 1990, research contributions were relatively limited and mainly focused on foundational statistical methods. The 1990s marked a turning point, possibly influenced by improvements in computational power and the growing complexity of power systems. During this time, early ML techniques and hybrid models for peak demand forecasting began to emerge. The 2010s saw a steady increase in research output, possibly influenced by the global push for sustainable energy and the increasing use of renewable energy sources. During this period, more ML techniques such as LSTM and CNN gained prominence, reflecting a shift towards models better suited to addressing the complexities of modern power systems in terms of accuracy and scalability. With these methodological and conceptual transformations, the variety of publication platforms has also increased. Early research was primarily published in engineering-focused journals, whereas more recent work has appeared increasingly in interdisciplinary publications covering energy economics, policy, and sustainability. This broader range of publications signifies a growing recognition that peak demand forecasting extends beyond technical modeling, playing an increasingly important role in strategic decision-making within power systems.

2.3. Unified framework for peak demand forecasting

Drawing from the analysis of the reviewed studies, a unified framework for peak demand forecasting was developed, as shown in Fig. 3. Peak demand forecasting begins by defining the geographical scope and time horizon. Next, key variables are identified and grouped into endogenous variables (e.g., historical peak loads) and exogenous variables (e.g., temperature changes, public events, and economic disruptions). Before model construction, many studies emphasize a dedicated data preparation step. In the context of peak demand forecasting, this step is shaped by the nature of the data. Load records from smart meters and feeders often contain gaps or sensor errors, which have been handled through imputation or multi-source fusion [20,21]. Normalization and scaling are frequently applied to stabilise training when variables differ in magnitude, especially in household and building-level studies [22,23]. Noise reduction and decomposition techniques such as empirical mode decomposition and wavelets are used to filter volatility and expose clearer structures in the load series [24,25]. Feature selection methods, most notably PCA, have been used to reduce dimensionality and highlight relevant weather or economic drivers [26,27]. Clustering methods such as SOM or K-means are applied for similar-day selection and to capture user groups with comparable behavior [28]. In rarer cases, data augmentation based on Extreme Value Theory (EVT) has been used to generate synthetic extremes [29], and robust regression frameworks explicitly account for anomalies [30]. Moreover, emerging smart meter and IoT sources further increase heterogeneity in data resolution and reliability, which calls for fusion and alignment strategies in data preprocessing.

Forecasting models are then developed using methods such as statistical approaches or ML techniques. Following this, the models are evaluated through a process of iterative adjustments to ensure reliable performance [31]. Finally, the process delivers the output, including the predicted peak value (maximum load) and/or the peak time. In the following subsections, each component of the framework will be discussed in detail.

2.3.1. Forecasting scope and horizon

Peak demand forecasting operates across multiple scopes, each tailored to the specific needs and objectives of various stakeholders in the electricity market (see Table 1). These scopes vary at national, regional, municipal, and local levels, each addressing unique challenges for peak demand management.

At the national level, peak demand forecasting supports large-scale energy policy-making and infrastructure development, addressing inter-regional energy transfers and balancing the national grid. At the regional level, the focus shifts to balancing grid operations within states, provinces, or utility regions, often aggregating demand across multiple municipalities or localities. The municipal level targets urban energy needs, supporting city or town-wide demand management, often relying on aggregated data from municipal grids or city-wide demand profiles. At the local level, forecasting focuses on smaller-scale entities, such as neighborhoods, microgrids, or distribution feeders, which require more granular data inputs (e.g., time-of-use data) to optimize operational efficiency.

It should be noted that the local and municipal levels are often collectively referred to as the ‘local’ level [32,33]. However, they differ in their focus within power systems. The local level emphasizes smaller, specific areas, capturing granular variations in peak demand, whereas the municipal level considers broader factors such as urban density and socio-economic activities that shape distinct load usage patterns [34]. This difference is important, as municipal areas function as intermediaries, translating regional energy strategies into localized actions [35]. By refining the geographical scope in peak demand forecasting, we can more accurately track trends in forecasting methods, explore spatial and socio-economic disparities across different geographic levels, and better capture the hierarchical structure of power systems.

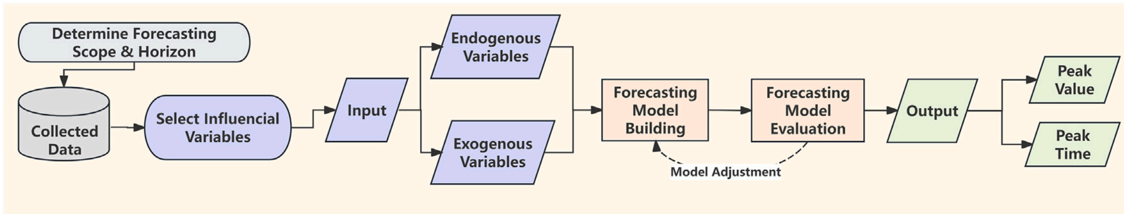

Fig. 3. Unified framework for peak demand forecasting.

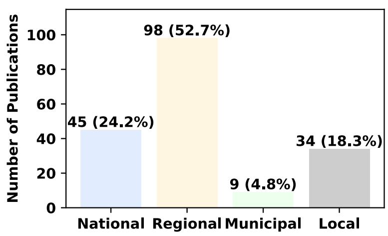

Fig. 4. Research distribution across forecasting scopes.

Table 3
Research distribution across forecasting scopes and stages.

Scope	Pre-ML	Transitioning	ML	Total
Regional	9	50	39	98
National	1	14	30	45
Local	1	4	29	34
Municipal	1	2	6	9

Fig. 4 illustrates the distribution of the reviewed studies across the four forecasting scopes. Overall, regional studies account for the largest share, 52.7 %, emphasizing their key role in grid management. While municipal work is scarce, 4.8 %, this reflects limited research at the city and town levels. However, these figures are cumulative totals for distinct stages, and the emphasis of research has shifted as the field moved from earlier phases into the machine learning stage. Table 3 shows the research distribution across different scopes in different stages. In the pre-ML and transitioning stages before 2011, regional studies far exceeded other scopes. Since the machine learning stage, the profile has shifted. Numbers for regional, national, and local studies are now similar, and municipal research totals nine, with six in this stage. Compared with earlier periods, national, local, and municipal research increased sharply while regional counts declined. This pattern indicates that the large number of regional studies reflects the focus of earlier stages rather than the current emphasis. The shift likely stems from improved access to granular data, together with stronger modeling toolkits. Although municipal and local studies remain underrepresented, the emergence of advanced ML techniques such as transfer learning, multitask learning, and multi-agent modeling offers promising avenues to improve peak demand forecasting at these levels.

In addition to its scope, peak demand forecasting also varies in its time horizon. Although the forecast horizon of the general load forecast is well known, there is no such summary for peak demand forecasting. Therefore, building on definitions commonly used in general load forecasting studies [14,36] and informed by patterns observed across the reviewed literature, we classify the time horizon of peak demand forecasting into the following categories:

- Short-term peak demand forecast (STPDF), to forecast peak demand from several hours to days ahead (days<7).
- Medium-term peak demand forecast (MTPDF), to forecast peak demand from weeks to months ahead (months<12).
- Long-term peak demand forecast (LTPDF), to forecast peak demand more than a year ahead.

Based on the reviewed literature, we present the primary focus for each forecasting horizon: 1) in STPDF, daily peak demand forecasting is most commonly utilized; 2) MTPDF is primarily concerned with weekly and monthly peak demand; and 3) LTPDF mainly focuses on annual peak demand forecasting.

2.3.2. *Influential variables in peak demand forecasting*

Accurate forecasting models rely on a combination of endogenous and exogenous variables that influence energy demand. Endogenous variables are internal factors that directly represent the behavior and dynamics of the system, while exogenous variables are external factors that impact energy demand but are not inherently part of the internal dynamics of the power system [37].

In the context of peak demand forecasting, *endogenous variables* primarily consist of historical peak load data. Specifically, peak load data from similar days is a commonly employed endogenous variable, as it allows forecasting algorithms to identify and utilize recurring patterns in demand, such as daily or seasonal fluctuations, which are relevant for making accurate predictions for different forecasting periods. For instance, Yu et al. [38] proposed algorithms that first identify similar days and then utilize the peak demands of these days as historical training data, thereby enhancing the model ability to predict future daily peaks accurately.

Unlike general load forecasting models, which predict overall load patterns over time using detailed, continuous datasets such as hourly data points for each day (i.e., 24 data points per day) or other granular intervals, peak demand forecasting models focus specifically on peak load values, such as the daily peak (i.e., a single data point per day) or other granular intervals. Peak demand forecasting models can also incorporate the peak load time (i.e., a data pair for each day). By focusing exclusively on peak load values and potentially the timing of these peaks, peak demand forecasting significantly reduces the number of data points. This reduction leads to higher computational efficiency. Amjady [39] compared the number of input features and computation time between the hourly load forecasting and the daily peak demand forecasting using the same historical data. Their statistical analysis revealed that only six input features were necessary for daily peak demand forecasting, whereas 171 input features were required for hourly load forecasting.

While endogenous variables such as historical peak load are indispensable, relying on them alone risks overlooking abrupt shocks driven by exogenous factors. For example, extreme weather, calendar events, and policy interventions can trigger sudden deviations that cannot be inferred solely from past peaks. Moreover, the interplay between endogenous and exogenous variables is rarely linear or unidirectional. Demand itself can influence exogenous drivers, such as through tariff changes or demand-response activation, creating feedback loops that complicate model design. Addressing these complexities requires models that can capture non-linear and bidirectional relationships, where machine learning methods such as neural networks or

Table 4
Paper distribution on the utilization of exogenous variables across different forecasting stages.

Stage	Weather	Calendar	Economic	Policy	None
Pre-ML	7 (77.8 %)	3 (33.3 %)	2 (22.2 %)	0 (0 %)	2 (22.2 %)
Transitioning to ML	61 (81.3 %)	28 (37.3 %)	15 (20.0 %)	0 (0 %)	5 (6.7 %)
ML	75 (74.3 %)	31 (30.7 %)	30 (29.7 %)	7 (6.9 %)	14 (13.9 %)

Table 5
Summary of the exogenous variables used for peak demand forecasting models. GDP: Gross Domestic Product; GNP: Gross National Product; CPI: Consumer Price Index; PV: Photovoltaic; HVAC: Heating, Ventilation, and Air Conditioning; DSM: Demand-Side Management; KPI: Key Performance Indicator.

Category	Commonly used variables	Other variables
Weather	Average temperature, maximum/minimum temperature, humidity, wind speed	Cooling/heating degree days, daily degree-days, heatwave patterns, precipitation, rainfall, soil temperature, atmospheric pressure, wind direction, solar radiation, luminosity, cloud height, visibility, seasonal patterns, temperature cycles, pressure, wind chill factor
Calendar	Day type classification (weekday/weekend), public holidays	Time of day, time of year, seasonality, religious events (e.g., Ramadan, Eid, Hajj), special socio-economic events, operational events, seasonal cyclic effects, trigonometric encodings, solar terms
Economic	GDP index, population growth/trends, market prices	GNP, CPI, general economic indicators, household demographics, appliance usage patterns, population and household data, industrial production, load factor, renewable energy capacities (e.g., PV capacity), socioeconomic trends
Policy	Tariff system changes, HVAC operation schedules	Policy goals, DSM participation rates, load diversity, time of use, building energy efficiency, grid quality parameters, KPIs

recurrent architectures have clear advantages over static regression frameworks.

Exogenous variables provide additional contextual information that enhances the accuracy and reliability of peak demand forecasting models. Based on the reviewed studies, the exogenous variables used are classified into four main groups: weather, calendar, economic, and policy. Table 4 highlights the evolving use of these exogenous variables in peak demand forecasting across three development stages. Weather variables are observed to consistently play a key role at every stage. Economic variables, on the other hand, show a marked increase in their influence during the ML phase. Calendar factors display a more complex trend, initially rising during the transition stage and then slightly decreasing during the ML stage. Policy variables are rarely used at all stages, with some studies choosing to exclude exogenous variables. This omission might signify difficulties in merging diverse variables or suggest that simpler models, which rely solely on endogenous variables, are more practical or effective based on the forecasting context and available data.

Table 5 summarizes the exogenous variables employed in the peak demand forecasting models, highlighting the commonly used variables identified in the reviewed studies. Among these, weather variables such as maximum and minimum temperature are important as they capture the variations in temperature that directly influence energy demand patterns. It should also be noted that some variables are internally related and can affect each other. For example, Contaxi et al. [40] noted that high relative humidity during months with apparent seasonality (e.g., summer or winter) can lead to increased demand for cooling or heating, which in turn impacts the accuracy of peak demand forecasts. To address this issue, the authors proposed quantifying relative humidity in terms of its effect on temperature fluctuations. By integrating this adjustment into the model, they were able to correct inaccurate input variables, thereby improving the forecasting accuracy for peak demand. A further limitation arises from reliance on single weather station measurements. Forecast accuracy can suffer when there is spatial misalignment between observation sites and demand centers, or when systematic biases are present in the instruments. Recent research addresses these challenges through the integration of multi-source meteorological datasets, including reanalysis products and satellite-derived indicators, which enhance the robustness of weather-driven peak demand forecasts.

Calendar variables have a significant influence in areas with rare special events and regular holidays. In the study by Saini and Soni [26], the influence of lunar calendar festivals in Egypt on peak load was considered, and the influence of Ramadan is quantified as a weight factor and input into the expert system. The prediction results showed that the models considering special festivals performed better than others. Moreover, load consumption on the weekends and holidays of commercial and industrial sectors is considerably different from that of working days, and the peak demand may not even occur in these sectors during non-working days. Therefore, some of the reviewed works separately modeled historical data based on these calendar factors to improve the forecast performance. Ramanathan et al. [41] trained models separately for each hour of a weekday and weekend day, respectively, which resulted in 48 independent models to predict the daily morning peak and the afternoon peak. This time-division modeling method distinguishes between working days and non-working days, which overcomes the defect of the traditional model in peak demand forecasting on weekends.

Socio-economic indicators such as Gross Domestic Product (GDP), household demographics, and market prices do more than provide long-term background trends, they also shape the structural composition of demand. Industrial growth, urbanization, and changes in household appliance ownership can alter both the magnitude and the timing of peak loads. Incorporating these drivers helps forecasting models remain sensitive to structural shifts in consumption patterns and improves their ability to anticipate emerging demand pressures.

Economic variables capture the broader socio-economic factors that influence electricity consumption patterns over medium to long-term horizons. The increasing incorporation of economic variables, especially in the ML-driven peak demand forecasting stage, demonstrates their importance in accounting for factors such as economic expansion, technological advancements, and shifts in consumer behavior. Commonly utilized economic indicators include the GDP index, population trends, and market prices, which provide information on overall economic health and demographic changes that affect load demand. Moreover, policy variables include regulatory and policy-driven factors that can significantly impact load consumption. For instance, Jang et al. [42] calculate the demand-saving impact of DSM efforts and integrate this into a probabilistic framework for LTPDF, effectively treating DSM as a key factor in mitigating future peak electricity loads. Although policy variables are less frequently integrated compared to other exogenous variables, they are essential to align forecasting models with government regulations and strategic energy initiatives.

2.3.3. Evaluation indicators for peak demand forecast

The evaluation indicators of peak demand forecasting models typically rely on both standard accuracy measures and specialized indicators designed for peak-specific contexts. These indicators build upon the formalization of peak demand as the maximum load observed over the specified horizon, reflecting both system stress and cost escalation under peak conditions. Commonly adopted metrics include *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), and R^2. Although these metrics are well established in general load forecasting [43], their implications in the context of peak demand forecasting merit further discussion.

General error metrics. MAE has been widely applied in early neural network studies [39,44], offering an intuitive measure of average error magnitude in megawatts, which is easily interpretable by operators. However, because it weights all deviations equally, it does not emphasize extreme errors. MSE and RMSE are more sensitive to large deviations and are often preferred when peak errors are operationally critical [45,46]. Yet, this quadratic penalization makes RMSE particularly sensitive to outliers, as shown in studies under atypical days or extreme weather conditions [47,48]. MAPE remains the most reported indicator across decades of research [8,49–51], due to its normalized form that allows comparison across regions and datasets. However, it becomes unstable when the actual load values are small (e.g., in feeder or household-level studies) [22,52], and it tends to exaggerate errors under non-normal distributions. R^2 is also reported in several regression-based and building-level works [53,54], which is the coefficient of determination, providing a measure of variance explained. While useful for model benchmarking, R^2 can be misleading in non-stationary contexts, since a high fit does not necessarily imply accurate peak capture. To avoid optimistic estimates, studies commonly adopt blocked or seasonal cross-validation and compare multiple metrics jointly (e.g., MAPE with RMSE and interval coverage), especially when data are noisy or imbalanced.

Peak-specific indicators. To better capture performance at extreme values, *Peak Absolute Percentage Error* (PAPE) was introduced in hybrid models [55], quantifying deviations only at the daily or seasonal maximum. Assuming that $\hat{y}$ is the predicted peak value, y is the actual peak value, and n is the number of training samples:

$$\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\} \tag{1}$$

$$y = \{y_1, y_2, \dots, y_n\} \tag{2}$$

PAPE [55] is defined as:

$$\text{PAPE} = \frac{100\ \%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{3}$$

The range of PAPE is $[0, +\infty)$. PAPE equal to 0 % represents a perfectly trained model, while PAPE greater than 100 % indicates an unacceptable model. While PAPE focuses attention on peak accuracy, it inherits the instability of MAPE under small denominators and does not provide any temporal dimension of error.

Temporal accuracy of peak demand forecasting models is often assessed through the *Hit Rate* (HR), which measures whether predicted peaks occur within a specified tolerance window [55,56]. Assuming that $\hat{t}$ is the predicted peak time, t is the actual peak time, n is the number of training samples:

$$\hat{t} = \{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_n\} \tag{4}$$

$$t = \{t_1, t_2, \dots, t_n\} \tag{5}$$

By using ϵ to represent the tolerance residual for the peak time and h as a flag to represent whether the predicted time falls within the tolerance interval $[t_i - \epsilon, t_i + \epsilon]$, then we have HR [55] defined as follows.

$$\text{HR} = \frac{100\ \%}{n} \sum_{i=1}^{n} h_i, \tag{6}$$

where

$$h_i = \begin{cases} 1 & \hat{t}_i \in [t_i - \epsilon, t_i + \epsilon] \\ 0 & otherwise \end{cases} \tag{7}$$

Peak time forecasting is usually measured by HR [55,57], which specifies a period before and after peak demand occurrence as the forecast error tolerance range. The prediction is considered correct as long as the predicted time falls within the tolerance interval. HR is particularly useful in operational contexts where anticipating the correct period is more valuable than predicting the exact timestamp. However, it has a binary nature: predictions just outside the tolerance interval are penalized equally to gross mispredictions, and it ignores the magnitude of the timing error.

Classification-based indicators. In some studies, peak demand forecasting is framed as a classification rather than a regression problem, for instance, identifying whether the next day will contain a critical peak event or whether a forecasted hour falls within a peak period [56,58–60]. In these cases, the evaluation is reported using accuracy, precision, recall, F1 score, or *Computed Area Under the Curve* (AUC) instead of continuous error metrics. Accuracy is commonly reported for daily peak occurrence [58], but these metrics are sensitive to class imbalance, since non-peak days vastly outnumber peak days in typical datasets. Thus, recall and AUC are often prioritised to ensure that peak events are not missed, even at the cost of higher false alarms.

Probabilistic and uncertainty-aware indicators. Unlike point-based measures such as MAE or MAPE, which only evaluate the average magnitude of forecast errors, probabilistic indicators assess the quality of the entire predictive distribution. They capture both the reliability and sharpness of interval or density forecasts, which is crucial in peak demand forecasting where operators must plan for uncertainty in extreme load values. The *Coverage Probability of Prediction Intervals* (PICP) measures the proportion of observed values that fall within a forecasted confidence interval [61]. A higher PICP indicates that uncertainty ranges are well calibrated, although excessively wide intervals may trivially achieve high coverage while providing little operational value. The *Continuous Ranked Probability Score* (CRPS) evaluates the distance between the cumulative distribution function of the forecast and the actual observation. It is analogous to MAE but for probabilistic forecasts: lower CRPS values indicate that the forecast distribution assigns higher probability mass close to the realized outcome [62,63]. Unlike PICP, CRPS rewards both calibration and sharpness simultaneously, making it one of the most widely used scoring rules for probabilistic load and peak demand forecasting. Other indicators include the *prediction interval normalized average width* and the *prediction interval coverage probability deviation*, which jointly evaluate the trade-off between interval sharpness and reliability [64]. These indicators collectively offer a clearer glimpse of model performance, indicating that even with a low point error like MAE, risk might be underestimated if intervals are consistently narrow.

Thus, for peak demand forecasts, probabilistic evaluation is crucial to prevent electricity shortfalls from under-prediction and extra costs from over-prediction.

An effective evaluation strategy for peak demand forecasting depends on aligning metric selection with the decision context, while also reflecting how evaluation practice has evolved in the literature. Across the reviewed studies, MAPE is the most frequently reported point metric (typically alongside RMSE/MAE), while classification metrics appear mainly when peaks are framed as events. Peak-specific indicators (PAPE, HR) are used in a minority of works to emphasize performance at extremes. Probabilistic and uncertainty-aware scores (e.g., PICP, CRPS) are reported in more recent ML papers and long-horizon settings, where calibrated intervals matter for planning. Rather than relying on any single indicator, studies increasingly report a small set (e.g., MAPE + RMSE + PICP/CRPS) to balance average accuracy, tail sensitivity, and uncertainty.

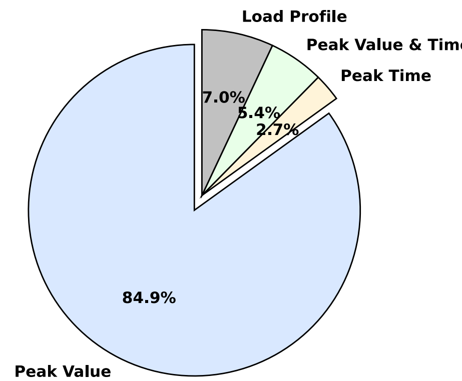

Fig. 5. The distribution of the peak demand outputs among reviewed studies.

2.3.4. *Outputs of peak demand forecasting*

Peak demand forecasting typically aims to predict a single value (e.g., peak load) or a pair of values (e.g., peak load and its time/date). Despite their significance, existing studies have not established a unified definition of the output of peak demand forecasting. By reviewing the literature, the following key types of output can be identified for peak demand forecasting:

(1) Peak time: Forecasting the time of peak demand.
(2) Peak value: Forecasting the peak demand value.
(3) Peak value & time: Forecasting both the peak demand value and its time.
(4) Load profile: Forecasting peak (or together with valley) values as intermediate results, serving as inputs to produce load profiles.

Each of these output types carries distinct practical relevance depending on the stakeholder. Peak time forecasts are particularly valuable for demand response program activation and load shifting. Forecasting the peak value helps utilities with generation scheduling and infrastructure stress testing. Forecasting both the peak value and time offers more granular control for system operators and facilitates dynamic pricing and operational planning [65]. Meanwhile, load profile forecasts are especially useful for long-term infrastructure investment, regulatory planning, and tariff design, making them more relevant to policymakers. Consumers also benefit indirectly, as accurate forecasting across these outputs contributes to improved grid reliability, fairer pricing mechanisms, and enhanced integration of distributed energy resources.

Fig. 5 illustrates the distribution of these forecast outputs across reviewed studies. It can be seen that the vast majority of studies focus on forecasting the peak value, while a smaller proportion considers forecasting the peak (and sometimes valley) value as an intermediate result to generate load profiles. Meanwhile, only a minimal fraction of studies focus exclusively on predicting the peak time. In addition, forecasting both the peak value and its time accounts for 5.4 % of the reviewed studies, which provide a more comprehensive forecast addressing not only the magnitude of the peak demand but also its temporal characteristics. Although less common, multi-output models that jointly predict peak value, timing, or load profiles can exploit correlations among outputs, offering stakeholders richer and more integrated insights for operational and market decision-making.

Beyond peak value and timing, outputs that capture duration and variability have important applications. Forecasting the persistence of high load periods informs system flexibility and reserve requirements, while variability measures are directly linked to ramping needs under renewable integration. Such outputs also support demand response, storage scheduling, and long-term planning based on load duration curves [21,29,66], highlighting their growing relevance in renewable and Distributed Energy Resources (DERs)-rich systems. In this review, we focus on peak value and timing, as these remain the dominant outputs in the literature and are most directly linked to system operation. While we do not provide a systematic quantification of the duration- and variability-aware outputs here, we revisit these outputs in the context of renewable integration and DERs, where ramping and persistence materially affect operations and planning.

2.4. *Primary trends and insights*

In this section, we provide a comprehensive summary of the literature on peak demand forecasting models by systematically mapping the studies to the previously defined unified framework, as shown in Table 6. This framework serves as a structured approach to categorize and compare the different methodologies and findings, highlighting key trends, models, and variables considered in peak demand forecasting. It should be noted that some of the reviewed studies employ hybrid models, rather than relying on a single method, and are not listed in Table 6. These studies will be discussed later in this section.

2.4.1. *Peak demand forecasting insights*

In this section, we provide a comprehensive summary of the literature on peak demand forecasting models by systematically mapping the studies to the previously defined unified framework, as shown in Table 6. This framework serves as a structured approach to categorize and compare the different methodologies and findings, highlighting key trends, models, and variables considered in peak demand forecasting.

The reviewed studies on peak demand forecasting reveal a complex interplay between the various components of the forecasting framework. A diverse range of methods has been employed, from traditional statistical approaches to ML techniques. Notably, the choice of methods appears to be closely related to the specific forecasting horizon. For STPDF, models are designed to capture immediate fluctuations in demand, relying heavily on high-frequency variables such as weather and calendar data. Commonly utilized methods include regression, stochastic time series, and ANN models. Ensemble learning methods, which combine multiple models, are also increasingly adopted to enhance robustness. The focus in this horizon covers all types of outputs, with a primary emphasis on predicting the peak value. Some attention is also given to the peak time or combined peak value with time, as well as to forecasting the peak demand as intermediates to assist in the prediction of load profiles. In this horizon, the data are typically stationary and abundant, allowing models to exploit fine-grained temporal patterns. Lightweight and interpretable methods often perform competitively, and the main challenge is modeling rapid demand variability.

In the context of MTPDF, the complexity of inputs and forecasting requirements increases. Models often incorporate economic variables alongside weather and calendar data to capture seasonal patterns and mid-term demand fluctuations. Stochastic time series models and ANNs

Table 6
Comprehensive summary of the reviewed literature mapped to the unified peak demand forecasting framework. W: Weather; C: Calendar; E: Economic; P: Policy; N: National; R: Regional; M: Municipal; L: Local; V: Peak Value; T: Peak Time; V + T: Peak Value & Time; LP: Load Profile.

Methods	Forecast horizon	Input variables				Geographic scope				Output				References
		W	C	E	P	N	R	M	L	V	T	V + T	LP	
	STPDF	✓	✓	✓		✓	✓		✓	✓				[41,67–75]
Regression	MTPDF	✓		✓			✓			✓				[76]
	LTPDF	✓		✓	✓	✓	✓		✓	✓				[53,54,66,77–80]
	STPDF	✓	✓			✓		✓	✓	✓				[39,81–85]
Stochastic time series	MTPDF	✓	✓	✓			✓			✓				[86–89]
	LTPDF	✓		✓		✓	✓			✓		✓		[90,91]
	STPDF			✓					✓	✓				[52]
Time series decomposition	MTPDF	✓				✓			✓	✓			✓	[92,93]
	LTPDF	✓	✓	✓			✓			✓				[62,94]
	STPDF													
Grey prediction	MTPDF													
	LTPDF	✓		✓		✓						✓		[57]
	STPDF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	[26,44,45,50,58,95–122]
ANN & variants	MTPDF	✓	✓	✓			✓			✓		✓		[123–125]
	LTPDF	✓		✓		✓	✓	✓		✓				[126–132]
	STPDF	✓					✓	✓			✓	✓		[133–135]
Fuzzy logic	MTPDF													
	LTPDF													
	STPDF	✓	✓				✓			✓				[136]
Genetic algorithm	MTPDF													
	LTPDF													
	STPDF	✓				✓				✓				[48]
Extreme value theory	MTPDF													
	LTPDF	✓		✓	✓		✓		✓	✓				[29,46]
	STPDF	✓						✓		✓				[137]
Bayesian networks	MTPDF			✓		✓				✓				[138]
	LTPDF	✓		✓		✓				✓				[139,140]
	STPDF	✓	✓	✓			✓		✓	✓				[141–146]
SVM/SVR	MTPDF													
	LTPDF													
	STPDF	✓	✓	✓			✓		✓	✓		✓		[20,147–151]
Ensemble learning	MTPDF													
	LTPDF													
	STPDF	✓	✓			✓					✓			[60]
CNN & variants	MTPDF													
	LTPDF													
	STPDF	✓	✓		✓		✓	✓	✓	✓	✓			[22,63,152–156]
RNN & variants	MTPDF	✓		✓		✓				✓				[157]
	LTPDF													

are commonly applied to MTPDF. Time series decomposition techniques also emerge as valuable tools for MTPDF, helping to separate trend and seasonality components. Most of these approaches focus on predicting peak values, whereas fewer studies expand their scope to forecast both peak value and time, and some concentrate on forecasting load profiles. Here, semi-stationary conditions are common, requiring models to handle both short-term fluctuations and seasonal effects. Feature decomposition and hybrid approaches become more important, and probabilistic outputs are often preferred to reflect medium-term uncertainty.

LTPDF is marked by increased uncertainty and fluctuation, and a wider range of input variables, including policy factors and long-term economic trends. Commonly utilized methods include regression, ANN, and grey prediction models. These long-term approaches often focus primarily on forecasting the peak value, with fewer studies addressing combined metrics to forecast both the peak value and its time. One critical challenge in LTPDF is the non-stationarity of input data, resulting from evolving policies, market structures, and technology adoption over time. To mitigate this, models must handle structural shifts and long-term uncertainty, often relying on advanced feature extraction methods, adaptable learning algorithms and interpretable modeling frameworks.

Table 6 also highlights notable differences in the geographic scopes addressed by forecasting studies. At the national level, STPDF and LTPDF are the most common focuses, addressing the need for both immediate operational adjustments and strategic, policy-driven planning. Methods like regression models and stochastic time series are frequently used, with ANNs and time series decomposition also playing a growing role in handling complex data and long-term trends. At the regional level, STPDF dominates, while MTPDF and LTPDF are explored with nearly equal attention. ANNs are widely used, alongside regression and ensemble models for robust and reliable predictions. The municipal level sees less exploration, with studies focusing mainly on STPDF and LTPDF. ANNs are the most commonly used methods in this scope, valued for their ability to handle localized data, whereas stochastic time series and Bayesian networks are occasionally applied to capture temporal patterns and uncertainties. The local scope is explored more frequently than the municipal scope, with a strong focus on STPDF to manage community-specific demand fluctuations. Regression, ANN, and Recurrent Neural Network (RNN) models are often used because of their ability to work

Table 7
Performance benchmarking of peak demand forecasting methods across three stages based on the reviewed studies.

Stage	Methods	General (MAPE)	Peak-specific	Classification	Probabilistic
Pre-ML	Regression (n = 4)	3–10 %	–	–	–
	Stochastic TS (n = 4)	4–8 %	–	–	–
Transitioning to ML	Regression (n = 7)	1.5–6 %	–	–	–
	Stochastic TS (n = 4)	1–6 %	–	–	80–90 % (PICP)
	Time series decomposition (n = 1)	1.8–3.0 %	–	–	–
	Grey prediction (n = 1)	~1.3 %	–	–	–
	ANN & variants (n = 25)	1–6 %	2–3 % (PAPE)	–	–
	Fuzzy logic (n = 2)	1–3 %	–	–	–
	Genetic algorithm (n = 1)	1.8–3 %	–	–	–
	Extreme value theory (n = 1)	–	–	–	Return levels (1–20 yr)
	SVM/SVR (n = 2)	0.8–2.5 %	–	–	–
ML	Regression (n = 5)	1.5–3.9 %	–	–	–
	Stochastic TS (n = 3)	1.0–6.0 %	–	–	–
	Time series decomposition (n = 3)	1.8–5.1 %	–	–	0.2–0.5 (CRPS)
	ANN & variants (n = 18)	0.5–5.6 %	2–5 % (PAPE), 85–93 % (HR)	85–92 % (Acc), ~95 % (Recall), ~0.97 (AUC)	–
	Fuzzy logic (n = 2)	0.6–2.3 %	–	–	–
	Extreme value theory (n = 2)	3.8–5.8 %	–	–	Return levels (2–10 yr)
	Bayesian/GP (n = 3)	1.5–5.7 %	–	–	90–100 % (PICP), 0.1–0.3 (CRPS)
	SVM/SVR (n = 4)	0.6–2 %[a]	–	–	85–95 % (PICP)
	Ensemble learning (n = 6)	1.6–5 %[b]	up to 95 % (HR)	90–96 % (Acc), >90 % (Recall)	90–95 % (PICP)
	CNN & variants (n = 1)	–	–	85–92 % (Acc)	–
	RNN & variants (n = 6)	1.0–5.7 %	1–3 % (PAPE)	–	–

Note: Reported ranges are aggregated from reviewed studies. General metrics use MAPE for comparability.SVM/SVR typically 0.6–2 %; one study reported up to ~16.8 % [146]. Ensemble learning typically 1.6–5 %; one study reported ~23 % MAPE [20].

with fine-grained sequential data. Across all scopes, STPDF is the most common, reflecting its importance for operational planning.

The choice of outputs further influences the selection of forecasting approaches. The peak value remains the most commonly forecasted output, given its essential role in operational planning. There is a growing interest in more complex outputs including forecasting both the peak value and time, and load profiles, which require flexible techniques like grey prediction, fuzzy logic, and ensemble learning methods. These outputs provide richer insights into demand dynamics but are currently less frequently addressed in the reviewed literature. Where reported, peak-specific measures such as PAPE and HR confirm notable gains in capturing the timing of peaks, while classification metrics appear in more recent work that frames peak prediction as an event detection problem, extending the evaluation beyond conventional error measures.

The interconnection among these factors reflects the complexity of peak demand forecasting, where various aspects of the forecasting framework must be considered together. Forecasting horizons define the scope of the analysis, which in turn influences the selection of input variables and the resolution of geographic scales. Similarly, the choice of outputs shapes the methodological approach, determining whether a model focuses on aggregated indicators like peak value or more complex metrics such as forecasting both the peak value and time. This highlights the essential need to adapt forecasting techniques to the unique combinations of different geographic regions, varying time horizons, and the specific objectives of diverse stakeholder groups. Building on the above comparison, we further carried out a detailed performance evaluation of peak demand forecasting method families across stages, as shown in Table 7, offering a clearer view of the evolution in peak demand forecasting practice.

In the pre-ML stage, forecasting methods were largely limited to traditional regression and stochastic time series approaches, yielding general error rates in the range of 3–10 % [67,69,76] and 4–8 % MAPE [158,159], respectively. These methods lacked tailored peak-specific evaluation and could not accommodate classification or probabilistic forecasts, limiting their applicability in handling uncertainty or rare peak events.

During the transitioning-to-ML stage, a broader set of techniques emerged. Regression and stochastic models were retained but improved, showing reduced MAPE ranges (down to 1–6 %) [39,41,61,70–72,78,79,88]. Notably, hybrid architectures began to incorporate decomposition (e.g., time-series and grey models), soft computing (fuzzy logic, ANN), and evolutionary techniques (e.g., genetic algorithms), improving model responsiveness to the nonlinearities of peak load behavior. ANN variants became especially prominent, with MAPE as low as 1 % [160,161], and PAPE between 2–3 % [39]. Fuzzy logic models showed low MAPE (1–3 %) [162,163], highlighting early efforts to incorporate interpretability and rule-based reasoning into peak load forecasting. Although still nascent, probabilistic forecasting emerged using stochastic and EVT-based methods, reporting meaningful indicators such as PICP and return level intervals for extreme peaks.

In the ML-driven stage, model sophistication and diversity significantly increased. ANN variants consistently demonstrated high performance, with general MAPE as low as 0.5 % [131] and peak-specific classification metrics reaching 95 % recall and 0.97 AUC [58]. Ensemble learning and deep neural architectures (RNNs, CNNs) enabled improved learning of temporal and spatial dependencies, achieving high classification accuracy (up to 96 %) [148,155] and HR for peak events (up to 95 %) [150]. Probabilistic forecasting became more robust, particularly through Bayesian models, Gaussian processes, and ensembles, with PICP ranging from 85–100 % [63,64,137] and CRPS scores as low as 0.1 [62,164], indicating reliable interval forecasts. Moreover, dimensionality reduction and decomposition strategies (e.g., wavelet, ICEEMDAN) were increasingly integrated into ML pipelines to handle high-dimensional data and preserve peak-relevant patterns.

Overall, the evolution reveals a clear trend: from deterministic models centered on error reduction in the early stages, to more holistic, adaptive, and uncertainty-aware models in the ML stage. The shift toward peak-specific, classification, and probabilistic metrics highlights a growing awareness that peak demand forecasting requires both accuracy and interpretability under uncertainty.

2.4.2. Hybrid peak demand forecasting insights

Expanding on the findings from the analysis of individual forecasting methods, Table 8 provides a summary of hybrid models used in the reviewed literature. In particular, techniques such as exponential smoothing, Kalman filtering, and expert systems, which were not included in the previous Table 6 of standalone methods, are present in hybrid models. Hybrid models integrate different methods at various

Table 8
A summary of hybrid models utilized across different stages, with the ML components highlighted using underlines.

Pre-ML Stage
Stochastic Time Series + Exponential Smoothing [165]
Regression + Time Series Decomposition [158]
Stochastic Time Series + Time Series Decomposition + Exponential Smoothing [166]
Regression + Expert Systems [159]
Transitioning to ML
Regression + Stochastic Time Series [99,167]
Fuzzy Logic + Expert Systems [162,168]
Regression + Time Series Decomposition [169–171]
ANN + Fuzzy Logic [172–175]
Stochastic Time Series + ANN [47]
Stochastic Time Series + Exponential Smoothing [176]
ANN + Ensemble Learning [160]
Regression + Fuzzy Logic [30,177,178]
Stochastic Time Series + Expert Systems [39]
Regression + Expert Systems [179]
ANN + SOM [27,180]
ANN + Genetic Algorithm [181]
Expert Systems + SOM + Fuzzy Logic [28]
Stochastic Time Series + Time Series Decomposition [24,25]
SVM/SVR + Time Series Decomposition [182]
Stochastic Time Series + Grey Prediction [183]
Stochastic Time Series + Fuzzy Logic [161]
Fuzzy Logic + Genetic Algorithm [163]
ML Stage
Stochastic Time Series + Time Series Decomposition [184]
ANN + Genetic Algorithm [185–189]
ANN + Genetic Algorithm + SVM/SVR [55,190]
Regression + Genetic Algorithm [191]
Regression + Time Series Decomposition [192,193]
ANN + Fuzzy Logic [194]
Regression + Stochastic Time Series [195]
Time Series Decomposition + Fuzzy Logic + Exponential Smoothing [196]
Regression + ANN [197–199]
Time Series Decomposition + SVM/SVR [51,200]
Regression + Bayesian Networks + SVM/SVR [201]
Time Series Decomposition + GRU [38]
Time Series Decomposition + Kalman Filtering [202]
Stochastic Time Series + SVM/SVR + ANN [203]
SVM/SVR + ANN [204]
Time Series Decomposition + Exponential Smoothing [205]
Regression + Extreme Value Theory [42]
Stochastic Time Series + LSTM [8]
ANN + LSTM [206]
CNN + LSTM [164]
Time Series Decomposition + LSTM [21]
Time Series Decomposition + CNN [7]
ANN + Ensemble Learning [207]
Time Series Decomposition + RNN [64]

stages of peak demand forecasting. In the table, we specifically highlight the ML components within these models to offer a clearer perspective on the role and impact of ML in advancing peak demand forecasting.

Hybrid methods aim to address the limitations of standalone techniques by leveraging their complementary strengths, enhancing the performance of peak demand forecasting models. It can be seen that the advancement of hybrid methods corresponds to the developmental trajectory of the three stages of peak demand forecasting, demonstrating a shift from traditional statistical methods to ML-driven models.

In the pre-ML stage, traditional statistical methods were the primary focus of research. Most studies explored combinations of statistical techniques to improve forecasting accuracy and address various modeling challenges.

Later, the transitioning stage marked a gradual integration of ML methods alongside traditional statistical approaches. During this phase, while most studies continued to combine statistical methods, some began exploring the potential of blending statistical and ML techniques. Examples include regression combined with fuzzy logic and stochastic time series combined with ANN, demonstrating how ML could complement existing statistical models. Hybrid ML methods have also started gaining traction, with the combination of ANN and fuzzy logic emerging as a popular approach. This stage highlights a shift in research methods, where statistical methods remained central, but ML methods were increasingly recognized for their potential to enhance peak demand forecasting models.

During the ML stage, ML methods became a key component of hybrid models. Hybrid models combining ANN with genetic algorithms were frequently used [208]. Moreover, models integrating statistical methods with ML methods remained popular, including combinations such as regression with ANN and time series decomposition with different ML methods such as CNN, RNN, LSTM, and Gated Recurrent Unit (GRU). However, hybrid models that combined only statistical techniques became relatively rare. The ML stage marked a growing emphasis on ML-driven hybrid models, reflecting a shift toward advanced ML techniques to address complex research problems in modern power systems.

The evolution of hybrid methods across the three stages shows both continuity and progression. Importantly, there is overlap across stages: some purely statistical hybrids continued to be used well into the ML stage, while physics-informed hybrids emerged earlier than expected in specific long-term applications. This overlap suggests that the categorisation into three stages is a simplification, but it is still useful to capture the broad trajectory of methodological development. Moreover, within the literature, three main strategies for hybrid model integration can be observed. Sequential hybrids detrend or decompose demand data before ML captures non-linear residuals. Parallel hybrids generate forecasts independently with statistical and ML models and combine them through averaging or ensembles. Physics-informed hybrids embed domain knowledge and physical constraints directly into ML training. Table 9 provides a summary of the hybrid models reviewed, mapped to the unified peak demand forecasting framework, including their distribution over horizons, the top two most used input combinations, outputs, and their reported error reduction range.

Building on this summary, clear patterns emerge regarding the benefits of each approach. Sequential hybrids dominate STPDF, where preprocessing improves robustness and yields error reductions of up to 45 % while preserving some interpretability. Parallel hybrids, more evenly spread across STPDF and MTPDF applications, deliver 10–30 % improvements by combining complementary error patterns, offering greater stability across feeders. Physics-informed hybrids, though fewer, are concentrated in MTPDF and LTPDF, embedding economic or DSM factors alongside weather and calendar drivers. These deliver the greatest gains and generalise better to unseen scenarios. Alongside these benefits, several challenges were noted across the reviewed studies. Sequential hybrids, while interpretable, add preprocessing complexity and depend heavily on decomposition quality. Parallel hybrids reduce transparency and require careful weight calibration to avoid bias. Physics-informed hybrids demand richer data and more complex model design.

These differences illustrate the trade-offs between complexity, interpretability, and accuracy. Sequential hybrids balance simplicity and robustness, parallel hybrids favour accuracy and stability at the cost of clarity, and physics-informed hybrids extend reliability for planning tasks but demand greater modeling effort, adding design complexity and interpretability challenges, highlighting the trade-offs faced in this area. The choice of hybrid strategy therefore aligns with forecasting needs: sequential for short-term operations, parallel for regional or feeder stability, and physics-informed for long-term capacity and policy planning. This progression also indicates that while ML components increasingly dominate hybrid models, traditional statistical methods remain complementary. They provide interpretable baselines, reduce

Table 9
Summary of the hybrid models in the reviewed literature mapped to the unified peak demand framework. W: Weather; C: Calendar; E: Economic; P: Policy; N: National; R: Regional; M: Municipal; L: Local; V: Peak Value; T: Peak Time; V + T: Peak Value & Time; LP: Load Profile.

Strategy	Typical setup	Forecast horizon	Input (top 2)	Geographic scope	Outputs	Improvement	#papers
Sequential	Decomposition or smoothing → ML ANN, SVR, fuzzy)	STPDF (18) MTPDF (5) LTPDF (3)	W + C (12) W (5)	R (11) N (7) L (5) M (3)	V (22) T (2) V + T (1)	20–45 %	26
Parallel	ARIMA/SARIMAX + ML (ANN, SVM) via averaging/ensembles	STPDF (13) MTPDF (6) LTPDF (3)	W + C (10) W + C + E (5)	R (7) N (6) L (5) M (4)	V (14) T (3) V + T (3) LP (2)	10–30 %	22
Physics-informed	ML with domain knowledge (expert systems, EVT, Kalman)	MTPDF (7) LTPDF (6) STPDF (5)	W + C + E (6) W + C + E + P (5)	N (6) R (6) L (4) M (2)	LP (8) V (7) T (2) V + T (1)	25–50 %	18

Note: The reported improvement ranges are calculated as the relative percentage reduction in error (e.g., MAPE, RMSE, MAE) compared with the baseline model(s) reported in each study. For example, if a baseline MAPE was 5 % and the hybrid method achieved 3 %, the improvement was recorded as $(5-3)/5 = 40\ \%$. The ranges shown summarize the minimum and maximum improvements observed across all reviewed studies within each strategy.

noise before ML processing, and in many cases remain as benchmarks or support modules, highlighting their lasting significance in peak demand forecasting.

3. Fundamental context: pre-ML stage

In the pre-ML stage, various statistical methods were employed for peak demand forecasting. Fig. 6 shows both a word cloud and an inclusive counting bar chart. The word cloud conveys the diversity of methods reported in the literature, while the bar chart reports their frequency using inclusive counting, where hybrid studies are counted under each of their constituent methods. Regression analysis and stochastic time series models were the most widely applied techniques. Decomposition methods and exponential smoothing were also popularly utilized to identify trends and seasonal patterns in demand. Peak demand forecasting models during the pre-ML stage provided interpretable solutions with low computational demands, aligning with the technological constraints in early power systems. This section will first examine these constraints and then discuss the development of forecasting methods designed to address these challenges.

3.1. Challenges in early power systems

Peak demand forecasting was an important yet complex task for maintaining reliability, optimizing operations, and supporting long-term planning in the early power systems. During this period, the challenges were closely tied to the operational and planning needs of power systems. Based on the reviewed studies, the key challenges in early systems can be summarized as follows.

A major challenge in peak demand forecasting was capturing the *variability and irregularity of peak demand patterns*, driven primarily by exogenous variables such as weather, economic activity, and social behavior. Extreme weather conditions, such as heatwaves or cold spells, often led to abrupt spikes in electricity consumption for cooling or heating purposes. For instance, El-Razaz and Al-Mohawes [81] forecasted weekly peak demand in Saudi Arabia and emphasized that temperature surges in the summer could cause abrupt rises in electricity usage, complicating the peak demand forecasting. Additionally, industrial activity and socio-economic growth introduced irregular trends that required careful modeling, especially for fast-growing economies [158,165]. In the early power systems, many peak demand forecasting models struggled to capture these abrupt changes, often requiring additional weather variables or binary event indicators to detect anomalies.

Peak demand forecasting also faces the challenge of integrating *seasonal and cyclical nature of the peak demand*, which has disproportionate impacts on peaks compared to overall demand. For instance, during holidays or special events, overall energy demand might rise moderately, but peak demand can experience a more significant increase, resulting in sharper spikes in energy use that must be accounted for in forecasting. Peak demand exhibits predictable daily, weekly, and seasonal cycles influenced by work schedules, holidays, and climatic changes. However, deviations caused by irregular events, such as public holidays, cultural celebrations, or unexpected industrial activity, pose additional complexities. Papalexopoulos [69] demonstrates the importance of incorporating calendar effects to address these anomalies, emphasizing the need for peak demand forecasting models capable of handling both regular cycles and unexpected disruptions.

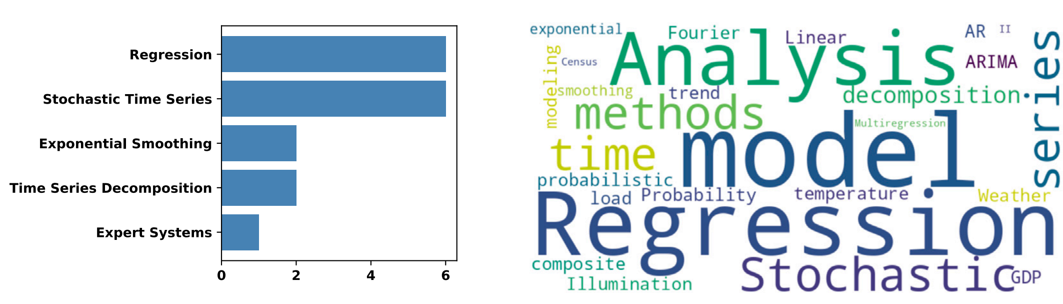

Fig. 6. Peak demand forecasting methods in the pre-ML stage. The bar chart (left) shows the quantitative frequency of methods based on inclusive counting, and the word cloud (right) provides an overview of the terminology used in the reviewed studies.

The *timing of peak demand* was another key focus. Accurate timing is essential for grid stability, as misaligned forecasts can lead to overloading or under-utilization of resources during peak periods. Grid operators need precise forecasts for peak times to prevent overloads and to optimize the utilization of available infrastructure. Hsu and Ho [168] highlighted the difficulty of synchronizing forecast peak times with real-world conditions, especially in regions with rapidly changing industrial or urban activity, where peak demand occurs with great variability.

Another challenge was the *limited availability of data* specifically tailored for peak demand forecasting. In early power systems, as energy demand continued to grow in rapidly developing regions, forecasting peak demand became increasingly challenging. Barakat and Eissa [158] emphasized the increased complexity of capturing peak demand data within the fast-growing regions. Moreover, high-resolution data, such as detailed hourly consumption patterns, were often unavailable in early power systems. While general demand forecasting models could rely on daily or weekly load profiles, peak demand forecasting required focusing on the highest load values within specific time intervals. The lack of such detailed data often limited the accuracy and effectiveness of peak demand forecasting models.

Finally, *operational and computational limitations* also restricted the complexity of the models used for peak demand forecasting. Early power systems required models that could provide actionable predictions to support operational decisions. While many peak demand forecasting models during this period provided interpretability and computational efficiency, these models often required manual adjustments to handle non-linearities and irregular events in the dataset [69,81,86,158].

3.2. Pre-ML peak demand forecasting methods

The challenges inherent in the early power systems gave rise to the use of statistical methods for peak demand forecasting. In the pre-ML stage, regression models were foundational for peak demand forecasting due to their simplicity and flexibility in capturing relationships between peak demand and explanatory variables. Many studies have applied regression models to forecast peak demand, incorporating various factors such as economic indicators, weather conditions, and special events. For example, Towill [76] utilized regression with GDP and temperature data to forecast peak demand for the CEGB North West Region in the UK. Stetson and Stark [68] applied regression to forecast the peak demand of rural residential customers in Kansas, USA, considering coincidental and non-coincidental peak demands. Moreover, Papalexopoulos [69] employed regression models incorporating holiday and temperature effects to perform STPDF. Despite their strengths, regression models often assumed static relationships between variables, limiting their adaptability to non-linear interactions and abrupt demand spikes [170].

Stochastic time series models were widely used to capture trends and seasonal patterns in peak demand. These models, such as ARIMA, autoregressive models, and moving average models, are effective for modeling and forecasting time-dependent data with inherent randomness. Based on the reviewed studies, these stochastic models were often combined with additional techniques to enhance forecasting accuracy. For example, El-Razaz and Al-Mohawes [81] utilized decomposition and autoregressive models to forecast weekly peak demand in fast-developing cities. Similarly, Elrazaz and Mazi [165] employed decomposition techniques in hybrid models combining ARIMA with exponential smoothing to forecast weekly peak demand. The stochastic time series models can effectively manage seasonal and cyclical variations but are constrained by assumptions of stationarity and require extensive parameter tuning, limiting their application in rapidly evolving power systems.

Time series decomposition methods have further enhanced peak demand forecasting by isolating various components such as trend, seasonal, cyclical, and irregular variations. For instance, early work by Gupta [86] applied Fourier transformations, a specific type of time series decomposition, to stabilize nonstationary series for monthly peak demand forecasting. Additionally, Barakat and Eissa [158] integrated decomposition with regression to address cyclic and irregular components in fast-growing regions. While decomposition methods improved the accuracy of forecasts by isolating different demand components, they often relied on aggregated data, limiting the granularity and responsiveness to short-term irregularities or real-time variations.

Exponential smoothing is another commonly used method in early power systems for peak demand forecasting, assigning exponentially decreasing weights to older data while emphasizing more recent observations. Variants such as single, double, and triple exponential smoothing (Holt-Winters method) address data with increasingly complex trends. For example, Barakat et al. [166] applied exponential smoothing for STPDF in fast-developing utilities. In the reviewed studies, exponential smoothing was frequently used in hybrid models to improve performance, often in combination with stochastic time series models. However, because exponential smoothing relies on manually selected smoothing coefficients, it may struggle to adapt effectively to highly volatile time series.

In the pre-ML peak demand forecasting, hybrid approaches combining multiple statistical techniques were also explored. Such as combining ARIMA models with decomposition and exponential smoothing [165], and the composite regression-decomposition models [158].

The methods outlined above formed the foundation for peak demand forecasting, addressing challenges such as variability, seasonality, and data limitations in early power systems. Table 10 summarizes the studies assigned to their respective methods, highlighting the challenges they tackled and their inherent limitations. Our review identifies a set of open questions that motivate subsequent developments. One concerns how regression frameworks can adapt to evolving non-linear relationships without relying on static specifications. Another relates to how stochastic time-series models can remain robust under structural breaks and rapid regime shifts while reducing the burden of manual parameter tuning. A further challenge is whether feature selection can be automated in ways that embed domain knowledge such as weather, tariffs, demand-side management, and policy, thereby reducing overfitting and improving transferability. An additional question concerns which

Table 10
A summary of peak demand forecasting methods in pre-ML stage.

Method	Description	Challenges Addressed	Limitations
Regression	Linear/parametric models with exogenous inputs.	Variability, weather, socio-economic drivers.	Static assumptions, weak for non-linearities or sudden spikes.
Decomposition	Trend–seasonality–noise separation.	Seasonal/holiday fluctuations.	Aggregated data only, poor for short-term irregularities.
Stochastic Time Series	ARIMA-type models exploiting autocorrelation.	Seasonal and cyclical patterns.	Assumed stationarity, limited adaptability.
Exponential Smoothing	Holt–Winters and variants.	Short-term trend smoothing.	Failed under abrupt shocks, inflexible for peaks.
Hybrid Models	Combined multiple statistical methods.	Integrated trend, seasonality, exogenous inputs.	Higher complexity and tuning effort.

decomposition strategies are best suited to capture high-resolution, short-lived irregularities rather than relying on aggregated seasonal patterns. Exponential-smoothing families also raise the issue of how to make them adaptive to volatility and peak episodes without repeated manual recalibration. A final question is how interpretable statistical hybrids should be designed so as to balance complexity, generalization, and transparency. These questions provide the foundation for the transition to the ML-era methods examined in the following sections.

4. Transitioning to ML-driven peak demand forecasting

The period between 1991 and 2010 represented a significant transition in peak demand forecasting methods, characterized by the evolution from traditional statistical approaches to ML-driven techniques. This era reflected the growing complexity of power systems and the need for more adaptable, scalable forecasting models to meet rising demand. Fig. 7 illustrates the methods adopted during the transition stage. It can be observed that key advancements during this period included the introduction of ANNs, fuzzy logic, and hybrid models that integrated statistical and ML methods [209]. While statistical methods like regression and ARIMA remained widely used, they were often combined with these more recent techniques to enhance model performance. Additionally, data-driven approaches became increasingly popular, incorporating tools such as clustering, optimization algorithms, and probabilistic forecasting approaches. These innovations marked a shift towards computational intelligence, addressing the rising challenges of modern power systems, and laying a foundation for future advancements.

This section is divided into two subsections: key advancements and breakthroughs. The key advancements subsection highlights methodological developments that have enhanced the performance of peak demand forecasting models over time. The breakthroughs subsection explores transformative innovations that have reshaped the field, introducing novel approaches and perspectives to address the complexities of peak demand forecasting.

4.1. Key advancements

4.1.1. Evolution of statistical methods

During this transitioning stage, a variety of statistical models from the pre-ML stage were enhanced to improve forecasting performance. Regression models remained foundational in peak demand forecasting, while significant advancements were made in stochastic time series modeling techniques. Barakat et al. [99] modeled seasonal load variations using Seasonal Autoregressive Integrated Moving Average (SARIMA), effectively capturing seasonal patterns. However, SARIMA presumes weak stationarity and linear dynamics, and its performance can deteriorate under structural breaks such as tariff changes, nonlinear weather–load interactions, or abrupt regime changes. To mitigate these limitations during this period, four extensions were explored: (i) non-stationary state-space formulations with time-varying parameters estimated via Kalman filtering [172]; (ii) ARIMA variants to capture volatility clustering [24]; (iii) decomposition-first pipelines (e.g., empirical mode decomposition) to isolate non-stationary components prior to residual modeling [94]; and (iv) SARIMAX and dynamic-regression structures that incorporate exogenous drivers directly into the linear framework [39,161,167].

Moreover, some complementary approaches, such as the Kalman filtering and grey prediction, also emerged to address uncertainty and data scarcity. Kalman filtering iteratively refines the forecasts through a prediction-correction process, as demonstrated in hybrid learning schemes for daily and weekly forecasts by Dash et al. [172]. Its main strengths are online error correction and robustness to missing or noisy measurements. The key limitations include sensitivity to mis-specified process and measurement noise, reliance on assumptions that are often linear and Gaussian, and limited ability to capture long-horizon trend shifts unless the state model is made explicitly time-varying. Meanwhile, grey prediction utilizes limited historical data to identify patterns and optimize models for peak demand forecasting. For example, Ran and Chaoyun [182] employed grey correlation theory to analyze the relationship between meteorological variables and the daily peak load, which is particularly useful in scenarios with insufficient historical data. Zhengyuan Yang et al. [57,183] both utilized hybrid grey models combined with ARIMA, improving accuracy for monthly and yearly forecasts, where the grey prediction is effective in identifying trends using minimal data, which complements the ability of ARIMA to manage non-stationary data. Grey prediction performs well in data-scarce and smoothly evolving settings, but it is sensitive to background-value choices and tends to degrade in volatile or regime-shifting systems. When combined with ARIMA, it can recover short-run dynamics while still retaining the advantages of small-sample performance.

4.1.2. Integration of ANNs

The emergence of ANNs, modeled after the structure and functioning of the human brain, marked a significant milestone in the evolution of peak demand forecasting. First introduced in the early 1990s, these models brought a new level of adaptability and capability to handle complex, non-linear relationships in forecasting scenarios. ANNs consist of interconnected artificial neurons organized in multiple layers to facilitate information processing and communication. A typical ANN architecture includes an input layer, one or more hidden layers, and an output layer. Each neuron, except those in the input layer, receives inputs from the previous layer through weighted connections. The aggregated input is then processed by an activation function to produce the output.

Studies from the early 1990s demonstrated the potential of ANNs in STPDF. For instance, Saeed Madani [98] utilized hourly temperature and load data from Seattle to develop an ANN model for forecasting daily and hourly peak demand, achieving significant accuracy improvements compared to traditional regression models. Hsu and Chen [50] utilized an ANN model to predict annual peak demand in Taiwan, which demonstrated the adaptability of the ANN model to non-linear dependencies in regional energy demand.

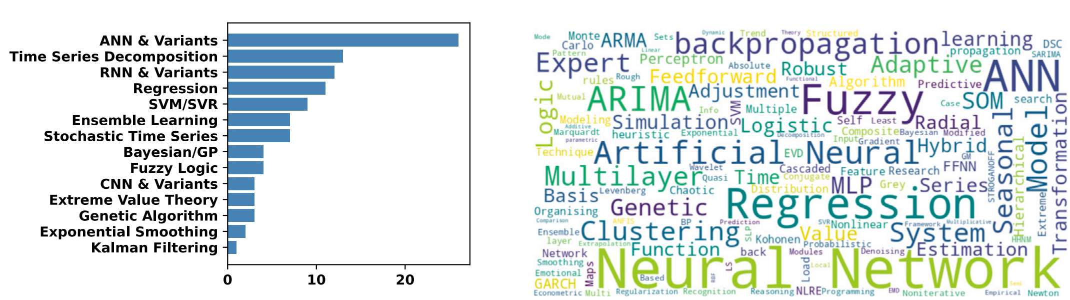

Fig. 7. Peak demand forecasting methods in the transitioning stage. The bar chart (left) shows the quantitative frequency of methods based on inclusive counting, and the word cloud (right) provides an overview of the terminology used in the reviewed studies.

Over the years, ANNs have evolved with the introduction of various architectures such as Multilayer Perceptrons (MLP), Radial Basis Function Networks (RBFN), and Self-Organizing Maps (SOM). For instance, MLPs have been applied using economic, demographic, and weather variables to predict regional peak demand, showcasing their adaptability in modeling diverse peak demand-related exogenous variables [50]. Moreover, RBFNs have been explored for computational efficiency. Nagasaka and Al Mamun [126] utilized RBFNs to predict daily peak demand for commercial buildings, demonstrating that RBFNs require less training time while maintaining high forecasting accuracy. In addition, SOMs have proven to be effective in clustering load profiles and isolating similar consumption patterns. For example, Hsu and Yang [95] employed the SOM to classify load patterns into day types, demonstrating their utility in identifying groups with similar hourly load profiles for STPDF. The flexibility and adaptability of ANNs and their variants make them well-suited to handle the variability in peak demand patterns. These advantages have established ANNs as a fundamental ML method for enhancing the accuracy and reliability of peak demand forecasting in evolving power systems.

Despite these advantages, ANN-based forecasting models face several limitations. Overfitting is a recurrent risk, particularly under sparse peak-event data, where models can memorise noise instead of capturing generalisable patterns [50]. Local minima also hinder convergence, especially in earlier training algorithms, although later optimization methods such as predictive backpropagation [100] and quasi-Newton schemes [26] have helped mitigate this challenge. Hyperparameter choices further play a critical role in ANN-based forecasting models: the number of hidden layers, neuron counts, and learning rates directly affect accuracy and generalizability, while activation functions such as sigmoid or ReLU determine the ability to model non-linear load–weather interactions. Regularisation methods, including Bayesian regularisation [110] and dropout, have proven essential for enhancing generalisation and reducing overfitting. Another important issue is robustness. ANN performance can deteriorate under changing data distributions, outliers, or missing values, which are common in peak demand datasets. Some studies have addressed this by applying rolling-origin validation [50], incorporating noise-resistant activation functions, detecting and filtering outliers before training [126], or using fuzzy pre-processing to handle incomplete inputs [173]. Overall, ANNs have clear potential for modeling variability in peak demand, but their performance in practice is strongly affected by the way architectures are designed and by how robustness is handled under noisy or incomplete data.

4.1.3. Introduction to fuzzy logic systems

During the transitioning stage, the incorporation of fuzzy logic systems into peak demand forecasting marked a significant advancement. Traditional deterministic models assume that the exogenous variables, such as temperature, have a fixed, predictable impact on electricity demand. However, in real-world scenarios, load data often include uncertainty and randomness, influenced by factors such as unexpected weather changes, human behavior, or industrial activity. Deterministic models struggle to capture these variations because they lack mechanisms to incorporate randomness or adapt to unexpected events. In contrast, fuzzy logic systems offer a more adaptable approach by utilizing linguistic variables and rule-based reasoning [210].

Early applications of fuzzy logic in peak demand forecasting frequently involved combining it with neural networks to develop hybrid models. For example, the fuzzy-neural network developed by [173] demonstrated a two-stage approach for short-term forecasting of peak and average loads. In the first stage, fuzzy rules were applied to preprocess inputs such as temperature, rainfall, wind speed, and seasonal variations, transforming these inputs into fuzzy linguistic variables. These processed inputs were then fed into an ANN, which captured the non-linear relationships between the fuzzy inputs and the forecasted load. This hybrid model proved capable of handling diverse conditions, including weekends, holidays, and extreme weather scenarios, achieving impressive forecasting accuracy with an average daily error below 0.7 %. Later, Kiartzis et al. [162] introduced a fuzzy logic expert system that further highlighted the adaptability of fuzzy logic. This system utilized a set of rules derived from historical data and expert knowledge to make predictions under uncertain conditions. By incorporating domain-specific heuristics, the fuzzy logic expert system significantly outperformed traditional statistical models in peak demand forecasting tasks.

Overall, the fuzzy logic systems and their hybrid models that combine with ANNs enhanced the adaptability and accuracy of peak demand forecasting, particularly in environments where demand was influenced by variable factors such as weather and operational changes. These models provided a powerful tool to manage dynamic energy demands by effectively handling ambiguity and integrating both quantitative and qualitative data, leading to more informed decision-making in power systems. However, a key challenge in fuzzy logic systems lies in the choice of membership functions and rule sets, which can be subjective and sensitive to the expertise of system designers. In peak demand forecasting, triangular or Gaussian membership functions have been commonly used, but studies show that performance varies with the functional form, requiring empirical calibration against historical data [162,173]. Rule-based reasoning also depends on the availability of reliable heuristics; incomplete or poorly specified rules can lead to inconsistent forecasts. Handling noisy or missing inputs is another limitation, since weather or economic variables are not always recorded at the needed resolution. Fuzzy systems partly address this by tolerating imprecision in input values, and hybrid fuzzy ANN or fuzzy expert designs have shown robustness when faced with incomplete information [30,177]. However, there is a continual risk that inaccurate or overly broad membership functions may decrease the accuracy of forecasts, emphasizing the importance of systematic validation.

4.1.4. Hybrid modeling approaches

During the transition stage, hybrid modeling approaches emerged as a response to the evolving complexity of power systems. On the one hand, the focus on hybrid statistical techniques is still popular, such as stochastic time series, regression, and decomposition methods. Table 11 provides a quantitative comparison of hybrid models with their baselines during the transition stage. It indicates that the statistical hybrids for peak demand forecasting generally delivered incremental improvements. These combinations aimed to enhance the accuracy and robustness of forecasts by leveraging the complementary strengths of statistical methods to handle seasonality, trends, and irregularities in demand data. For instance, Haida and Muto [49] utilized regression-based forecasting with transformations to capture non-linear relationships and seasonal fluctuations, particularly during transitional periods. Moreover, exponential smoothing has been commonly utilized to combine with stochastic time series [176] or decomposition methods [24], stabilising short-term fluctuations and improving the forecasting of volatile demand patterns. These statistical combinations formed the foundation of hybrid approaches, addressing many limitations of standalone statistical models but remaining constrained in their ability to manage non-linearity and uncertainty.

On the other hand, a notable trend in the transitioning stage is the combination of traditional statistical methods with emerging ML approaches. These combinations were aimed at overcoming the static assumptions of statistical methods and enhancing the adaptability of the forecasting models to non-linear and uncertain peak demand patterns. As shown in Table 11, these mixed hybrids proved considerably more effective. By incorporating ANN, expert systems, fuzzy logic, and grey prediction into time-series models, studies consistently reported stronger results than either statistical or ML baselines. Regression models were often integrated with ML techniques like fuzzy logic [30,177], expert systems [179], and time series decomposition [169,170], allowing for improved handling of uncertainty and non-linear patterns while

Table 11
Comparison of hybrid models with baseline models in the reviewed peak demand forecasting studies in the transitioning stage.

Study	Hybrid model	Baselines	Hybrid vs baseline MAPE (Improvement)
Hor et al. [24]	Stochastic TS + Decomposition	ARIMA	2.16 % vs 3–4 % (↓30–46 %)
Nazarko and Zalewski [177]	Regression + Fuzzy Logic	Regression	2.5–3.0 % vs 4–5 % (↓40–50 %)
n et al. [30]	Regression + Fuzzy Logic	Regression	0.44 % vs 0.84 % (↓48 %)
Kandil et al. [179]	Regression + Expert System	Regression;	2.6 % vs 4–5 % (↓40–48 %) [Regression]
		Box–Jenkins	3.4 % vs 6 % (↓43 %) [Box–Jenkins]
Haida and Muto [169]	Regression + Decomposition	Regression	Annual: 1.50 % vs 1.76 % (↓15 %)
			Spring/Fall: 1.43–1.68 % vs 1.93–2.16 % (↓20–30 %)
Haida et al. [170]	Regression + Decomposition	Regression	1.29 % vs 2–3 % (↓36–57 %)
Choi et al. [47]	Stochastic TS + ANN	Holt-Winters;	1.3–1.5 % vs 15–100 % (↓>90 %)
		SARIMA	
Fadhilah et al. [161]	Stochastic TS + Regression + Fuzzy Logic (ANFIS)	AR(2); RegARMA	1.47 % vs 1.27–3.45 % (↓0–57 %) [AR(2)]
			1.47 % vs 1.9–2.0 % (↓23 %) [RegARMA]
Zhengyuan [183]	Stochastic TS + Grey Prediction	GM(1,1); ARIMA	2.20 % vs 2.66 % (↓17 %) [GM(1,1)]
			2.20 % vs 4.72 % (↓53 %) [ARIMA]
Amjady [39]	Stochastic TS + Expert System	ARIMA; ANN	1.15–1.48 % vs 1.9–2.4 % (↓22–40 %) [ARIMA]
			1.15–1.48 % vs 1.2–1.5 % (↓0–20 %) [ANN]
Ran and Chaoyun [182]	Wavelet Denoising + SVM	SVM	2.6 % vs 3.5 % (↓26 %)
Dongxiao et al. [25]	Stochastic TS + Decomposition	ARIMA	2.0 % vs 2.5 % (↓20 %)
Gavrilas et al. [181]	ANN + Genetic Algorithm	ANN	11 % vs 10 % (↓9 %)
Amin-Naseri and Soroush [27]	SOM + ANN	ANN; Regression	1.83 % vs 2.02 % (↓9 %) [ANN]
			1.83 % vs 2.25–2.54 % (↓19–28 %) [Regression]
Wang [28]	Expert System + SOM + Fuzzy Logic	ANN	0.05 % vs 2–3 % (↓97–98 %)

Note: Metrics as reported in the original studies. Average absolute error (%) and average error (%) are equivalent to MAPE.

maintaining the interpretability of the peak demand forecasting models. Stochastic time series models were paired with ML approaches such as ANN [47], fuzzy logic [161], grey prediction [183], and expert systems [39], effectively managing time-dependent variations and benefiting from the pattern-recognition capabilities of the ML. Time series decomposition emerged as a common bridge, separating demand data into components such as trends and seasonality before applying ML methods like support vector machines [182] and ANN [25] to enhance the model performance. For example, Ghomi et al. [109] employed Fourier transformations to extract seasonal components, which were then fed into neural networks for STPDF, effectively capturing complex seasonal variations in the peak demand data. Decomposition proved especially useful in this period as it improved performance by allowing models to better capture seasonality and filter out noise. This highlighted the importance of preprocessing in the peak demand forecasting framework, and demonstrated the complementary roles of the two methods, where statistical approaches offered structure and clarity, and ML introduced flexibility and enhanced pattern recognition.

Moreover, there was a growing trend in developing advanced hybrid models that combined multiple ML techniques to further enhance performance. For instance, Gavrilas et al. [181] integrated ANN with genetic algorithms to optimize model parameters, while Amin-Naseri and Soroush [27] combined ANN with SOM to improve clustering and pattern recognition in peak demand data. Additionally, hybrid models incorporating fuzzy logic with genetic algorithms [163] or expert systems with SOM and fuzzy logic [28] demonstrated the potential of combining diverse ML methods to tackle the complexities of evolving power systems. The quantitative comparison in Table 11 also shows that the benefits of ML hybrids varied. Optimization-based hybrids produced only marginal improvements, while hybrids using clustering or fuzzy reasoning delivered the most substantial accuracy improvements. This suggests that hybrids added the most value when they introduced new ways of representing or structuring load patterns, rather than only tuning model parameters.

4.2. Breakthroughs

4.2.1. Feature engineering and selection

As power systems evolve, peak demand forecasting requires well-designed approaches to select and structure model inputs, effectively addressing the challenges of increasingly variable and extreme energy demand patterns. Unlike general forecasting methods, which often prioritize average trends under stable conditions, peak demand forecasting models focus on identifying and highlighting the key variables responsible for sudden and significant fluctuations in electricity consumption. This emphasis on targeted variables highlights the importance of feature engineering, which refines input data to improve model accuracy. Moreover, during the transition period, weather-related factors remain important for forecasting, but the changing energy landscape also requires incorporating a wider range of socio-economic variables.

Techniques such as Principal Component Analysis (PCA) have significantly enhanced the feature engineering process by identifying and emphasizing the most critical variables influencing peak loads, while effectively reducing noise and minimizing the impact of less relevant factors. For instance, Saini and Soni [26] employed PCA to preprocess weather and load-related variables, effectively minimizing redundancy in the dataset. This approach distilled 11 principal factors from an initial set of 28 variables, achieving a 96 % reduction in data dimensionality. The refined dataset not only facilitated more efficient neural network training but also improved interpretability by highlighting the weather variables most strongly associated with daily peak loads. Later, Amin-Naseri and Soroush [27] utilized PCA for dimensionality reduction in a feedforward neural network framework, enabling the model to concentrate on the most relevant variables. While PCA is effective for linear dimensionality reduction, it may not fully capture non-linear relationships in high-dimensional load data. Alternative techniques, such as t-distributed stochastic neighbor embedding, autoencoders, and Isomap, have been explored in related domains for their ability to preserve complex local structures or uncover latent representations. Future comparative studies could further evaluate the trade-offs among these approaches in terms of interpretability, robustness, and forecasting performance in peak demand applications.

With the transition of power systems towards accommodating increased economic activities and urban development, capturing a broader range of variables became essential during this stage. For example, Belzer and Kellogg [46] demonstrated the importance of integrating economic indicators such as GDP growth, which indirectly influence peak loads through changes in industrial and commercial energy use. Similarly, Saini and Soni [26] included economic factors to reflect the broader socio-economic context affecting electricity consumption patterns. Moreover, as energy markets became more competitive and tariff structures more complex, understanding the interplay between pricing mechanisms and consumer behavior became vital for accurate peak demand forecasting. Lee et al. [163] emphasized the role of electricity tariff structures in shaping peak demand, integrating tariff-related variables to better model the impact of economic policies on energy consumption patterns. Addressing the challenges posed by high-dimensional input spaces has become especially important as forecasting models incorporate growing numbers of features related to socio-economic, behavioral, and meteorological variables. Feature engineering techniques such as PCA, combined with targeted selection strategies, mitigate this issue by reducing noise and collinearity. Further, sparse representation techniques, such as Lasso-based selection, may support model generalization by enforcing parsimony, especially in high-dimensional settings. Manifold learning methods, including locally linear embedding, also offer potential to uncover intrinsic low-dimensional structures in complex load data, although their integration into operational forecasting remains limited.

4.2.2. Advancements in optimization algorithms

The integration of advanced optimization techniques with traditional and ML-driven peak demand forecasting methods has enabled models during the transition stage to be more accurate and robust.

One notable advancement is the application of evolutionary algorithms such as Genetic Algorithms (GA). For instance, Kato et al. [136] introduced a GA-based model, which utilized tree structures and the minimum description length principle to effectively model daily peak loads. This approach adeptly captured fluctuations in weather conditions and daily load patterns, resulting in notable improvements in the accuracy of forecasts. Gavrilas et al. [181] further combined ANN with genetic programming for peak demand forecasting in distribution systems. This hybrid approach used symbolic regression and correlation filtering to generate analytical expressions, matching the predictive accuracy of neural networks while improving interpretability and structural flexibility. GA became the optimiser of choice in peak demand forecasting because it handles mixed-variable optimization problems effectively, such as tuning ANN weights or fuzzy membership functions. Its global search ability makes it less prone to local minima, which is valuable in complex hybrid models. Fuzzy–GA hybrids provide a clear example of this trend. Lee et al. [163] demonstrated how GA can automate the tuning of fuzzy membership functions and rule weights, reducing the subjectivity of fuzzy design and adapting the system to changing peak demand conditions. This integration improved accuracy in uncertain scenarios but introduced added computational cost and reduced transparency, since both the fuzzy rule base and the optimization layer obscure interpretability. As such, fuzzy–GA hybrids highlight the trade-off between predictive power and explainability, a recurring theme in operational forecasting applications.

Although GA dominates the reviewed studies, this reflects both its early adoption and its suitability for mixed-variable optimization rather than inherent superiority. Other metaheuristics, such as particle swarm optimization and ant colony optimization, have been reported in related energy forecasting tasks [211–213], where they offered faster convergence or required fewer parameters to tune. Subsequent research could benefit from examining these options to enhance computational efficiency or robustness in peak demand forecasting.

Optimization techniques have also been important in training complex neural network architectures. Morioka et al. [100] introduced predictive backpropagation for next-day peak load forecasting, effectively handling dynamic time series data through incremental learning. Saini and Soni [26] employed Levenberg-Marquardt and quasi-Newton methods to accelerate convergence and improve accuracy, while Saini [110] utilized Bayesian regularization to mitigate overfitting and enhance generalization. Moreover, structural advancements such as hierarchical and cascaded neural networks complement these optimization trends. Carpinteiro et al. [180] proposed a hybrid hierarchical model combining self-organizing maps and single-layer perceptrons for improved LTPDF. Broadwater and Sargent [70] designed cascaded neural networks that leverage multi-layered outputs, reducing errors in short-term forecasts.

Robust regression techniques and least squares methods have also been employed to enhance the resilience of forecasting models against outliers and anomalies in load data. Haida et al. [170] introduced noniterative least absolute value estimation for LTPDF, effectively rejecting outliers and ensuring more stable and accurate annual peak load predictions even in the presence of anomalous data points. Moreover, Jin et al. [30] combined robust regression with fuzzy clustering to manage outliers and classify solar terms in daily peak load data, enhancing the model ability to handle extreme weather conditions and irregular load patterns. In the studies we reviewed, robust regression most often appeared in the form of least-absolute-value or quantile models. This has been a pragmatic choice for peak-load applications, as it accommodates asymmetric and heavy-tailed errors that arise during heatwaves or demand-side management events, and it can be implemented through linear-programming routines with stable run times. Huber-type M-estimators [214] can also be considered in peak demand forecasting, particularly for rolling feeder operations where anomalies are mild but frequent, although they are less commonly reported in the peak-demand literature because their down-weighting mechanism is less effective for extreme tails. The choice of robust regression therefore depends both on the error profile, whether dominated by tail risk or by mild contamination, and on the required update cadence, distinguishing between periodic planning and frequent online updates.

Support Vector Machine (SVM) and kernel-based optimization methods have also been explored for their capacity to handle non-linear relationships and high-dimensional data in peak demand forecasting. Wang et al. [141] applied least squares SVM for STPDF, leveraging SVM to model complex dependencies and achieve precise predictions even during special periods with extreme load consumption patterns. El-Attar et al. [142] combined Support Vector Regression (SVR) with a local prediction framework and correlation dimension-based time series reconstruction, effectively denoising load data using wavelet transformations and capturing intricate temporal dependencies for highly accurate peak demand forecasting.

4.2.3. Probabilistic forecasting approaches

Probabilistic forecasting approaches have gained substantial traction in peak demand forecasting due to their ability to quantify uncertainty and provide a range of possible peak demand forecasts. These methods are particularly valuable for risk management, capacity planning, and decision-making in increasingly complex and variable power systems.

Extreme Value Theory has played a key role in modeling the statistical behavior of extreme load events, which are essential for LTPDF. Belzer and Kellogg [46] incorporated Monte Carlo simulation with extreme value distribution to forecast long-term peak demand. Probabilistic models have also been developed to provide comprehensive forecasts that cover various sources of uncertainty in peak demand. Hor et al. [24] combined ARIMA with generalized autoregressive conditional heteroskedasticity models to conduct LTPDF. This hybrid approach enabled risk-based quantile estimation for load volatility, providing probabilistic forecasts that captured the variability and potential extremes in peak load values. Moreover, Mcsharry et al. [61] utilized stochastic time series models combined with probabilistic forecasting techniques to simulate weather variables and predict peak load magnitudes and timings, generating a range of possible peak demand outcomes that aid in assessing system reliability and formulating contingency plans.

While Bayesian networks offer a robust framework for probabilistic forecasting by modeling dependencies among various influencing factors, their application in peak demand forecasting was limited during the transitioning stage. The foundational principles of Bayesian networks were explored towards the end of this period, setting the stage for their broader adoption in subsequent years. Saini [110] introduced Bayesian regularization in ANN to incorporate uncertainty in peak demand forecasting. By leveraging Bayesian principles, the model was able to better handle parameter uncertainty and provide more reliable probabilistic forecasts, laying the groundwork for more efficient probabilistic modeling techniques in future research. Moreover, density forecasting techniques have further advanced probabilistic forecasting by providing comprehensive probabilistic insights into peak demand. Hyndman and Fan [79] utilized semi-parametric additive models combined with simulation techniques to forecast the density distribution of the long-term peak demand, incorporating temperature simulations and residual bootstrapping to provide detailed insights into the fluctuations of peak demand.

In addition to Bayesian networks, other advanced probabilistic methods such as Bayesian neural networks and Gaussian processes began to receive attention during this period for their capacity to represent parameter uncertainty and non-linear relationships. Bayesian neural networks incorporate prior distributions over weights and produce predictive distributions rather than point estimates, which are particularly valuable when data are sparse or noisy, a common situation in peak demand scenarios [110]. Gaussian processes, on the other hand, offer closed-form posterior distributions and uncertainty bands, making them highly interpretable and suitable for long-term planning. However, these methods also posed practical challenges, including high computational costs and difficulties with scaling to high-dimensional inputs, which constrained their widespread use during the transition stage.

5. Contemporary trends: ML-driven peak demand forecasting

The emergence of ML techniques has revolutionized peak demand forecasting, marking a new era characterized by the ability to model complex patterns and interactions within energy consumption data. A key enabler of this transformation is the availability of large, high-resolution datasets generated by the smart meters, which provide granular insights essential for accurate modeling and real-time decision-making [215]. In the ML stage, peak demand forecasting models leverage these datasets, enhanced computational power, and robust algorithms to achieve better forecasting performance. As illustrated in Fig. 8, it is also worth noting that some statistical methods, such as regression and stochastic time series models, remain popular in this era. They are especially useful when samples are short or when strong seasonal and calendar structures dominate. These models yield clear coefficients, often serving as the structural core in hybrid systems. For example, decomposition or state-space blocks address trend and seasonality, whereas a non-linear learner models the residual [21,202]. The non-linear component is typically modeled by ML. From a theoretical perspective, ML algorithms enhance forecasting accuracy by capturing non-linear relationships and variable interactions that classical models cannot easily represent. Neural networks approximate arbitrary non-linear functions through layered compositions of activation functions, with mechanisms such as dropout and weight regularisation preventing overfitting and improving generalisation [216,217]. Kernel-based methods like SVM/SVR project input data into high-dimensional feature spaces, making non-linear interactions linearly separable [218]. Ensemble methods reduce variance and bias by combining diverse learners, thereby stabilising forecasts across heterogeneous demand conditions [219,220]. These approaches effectively capture intricate dependencies within increasingly diverse and complex data sources, including those from renewable energy, active DSM, and decentralized generation. Methods in this era also emphasize probabilistic and explainable Artificial Intelligence (AI) models that address uncertainty and interpretability. Moreover, advanced feature engineering and transformation techniques have also enhanced the ML-driven peak demand forecasting to optimize the input representation for ML models.

Based on the above observations, it is evident that ML-driven methods have emerged as an increasingly important approach for modern peak demand forecasting. The growing availability of high-resolution datasets and advancements in computational power have further expanded the opportunities for leveraging these methods in this domain. To provide a comprehensive understanding of the evolution and significance of ML-driven methods in peak demand forecasting, this section offers an in-depth exploration of the most commonly used ML techniques in the field, highlighting their technical intricacies, applications across various contexts, and an assessment of their performance. By exploring the application contexts and evaluating these methods, this section aims to shed light on their transformative potential and provide insights into their contributions to overcoming the challenges of peak demand forecasting in modern power systems.

5.1. Techniques and models

5.1.1. Deep learning architectures

As a subset of ML, Deep Learning (DL) architectures leverage deep neural networks to automatically extract and represent complex, high-dimensional patterns. These architectures have significantly advanced peak demand forecasting by effectively modeling the intricate temporal and spatial dependencies inherent in electrical load data. By utilizing multiple layers of processing, DL architectures capture complex patterns and interactions, thereby improving the performance of peak demand forecasting models across diverse applications and contexts. As summarized earlier in Table 7, which benchmarks forecasting methods across stages, DL models generally achieve lower forecasting errors (0.5–5.7 %) than classical statistical approaches such as ARIMA or SARIMA (typically 3–10 %). Based on the reviewed studies, key DL architectures employed in peak demand forecasting are RNNs, CNNs, and hybrid models that integrate multiple DL techniques.

RNNs are particularly well-suited for sequential data processing, making them a powerful tool for peak demand forecasting. While traditional RNNs are good at capturing short-term dependencies, they face challenges with long-term dependencies due to vanishing gradients. Advanced variants such as LSTMs and GRUs address these limitations

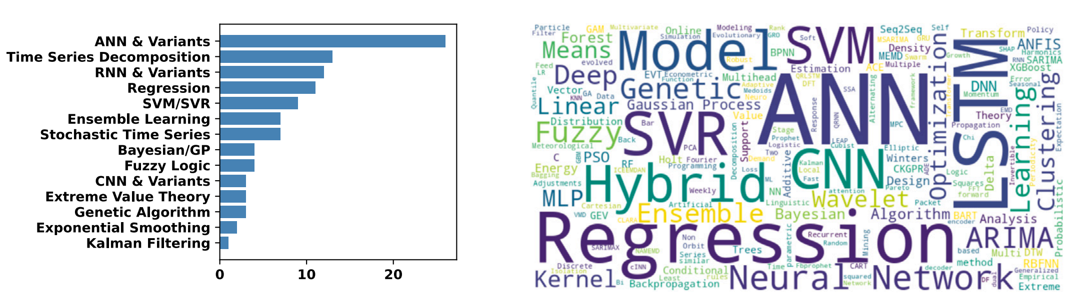

Fig. 8. Peak demand forecasting methods in the ML stage. The bar chart (left) shows the quantitative frequency of methods based on inclusive counting, while the word cloud (right) provides an intuitive overview of the terminology used in the reviewed studies.

by introducing mechanisms that enhance memory retention and selective information flow. LSTMs, with their memory cells and gating mechanisms (input, forget, and output gates), are adept at modeling long-term dependencies and sudden spikes in load data. These attributes make LSTMs invaluable for forecasting peak demand events characterized by abrupt changes. For instance, Mughees et al. [155] demonstrated the efficacy of a deep sequence-to-sequence bidirectional LSTM model for day-ahead peak demand forecasting, which utilized bidirectional processing of input sequences to capture comprehensive temporal patterns. GRUs, a streamlined alternative to LSTMs, reduce computational complexity by employing fewer parameters while maintaining comparable performance. Yu et al. [38] proposed a GRU model combined with Dynamic Time Warping (DTW) to enhance daily peak demand forecasting, achieving improved pattern recognition and reduced prediction errors. Across the reviewed studies, LSTM and GRU architectures were often trained using rolling-origin validation and tuned with respect to forecast horizon, enabling them to generalise under shifting load conditions. Compared to traditional RNNs, LSTM/GRU variants consistently improved robustness against noise, missing values, and distributional drift, particularly when combined with pre-processing or hybrid feature extraction.

CNNs are good at identifying patterns and extracting spatial features from multidimensional data, such as load curves. Their hierarchical learning capability enables the detection of detailed load patterns that may indicate or coincide with peak demand events. CNNs are particularly effective in scenarios requiring the decomposition and analysis of load data. For example, Liu and Brown [60] applied CNN models for predicting the timing of coincident daily peak loads, leveraging wavelet decomposition to enhance the extraction of temporal patterns and achieve greater forecast precision.

In peak demand forecasting, CNNs may fail to capture long-range temporal structures, while RNNs are prone to vanishing gradients. Hybrid CNN–LSTM architectures and dilated temporal convolutions have been proposed to address these limitations, as they combine local pattern extraction with extended receptive fields. In addition, residual and skip connections have been shown to improve stability during training. Zhang et al. [164] proposed a CNN-LSTM hybrid model with an encoder-decoder structure for LTPDF, demonstrating the synergy between CNN-based feature extraction and LSTM-based temporal learning. On the other hand, Liu and Brown [56] investigated the application of CNN and LSTM models to predict the occurrence time of daily peak loads. Their research revealed that LSTMs demonstrate proficiency in capturing temporal relationships, whereas CNNs are adept at feature extraction from historical datasets.

Recent advancements include models that incorporate even more complex architectures. Deng et al. [148] introduced CNN-Transformer hybrid models, which enhance adaptability to extreme weather conditions by integrating weather-dependent factors into load predictions. Additionally, He et al. [63] employed Conditional Invertible Neural Networks (cINN) in conjunction with LSTM variants to include uncertainty quantification, offering a probabilistic approach to peak demand forecasting.

5.1.2. Ensemble learning techniques

In the modern power systems, ensemble learning has emerged as a robust approach to enhance the accuracy and reliability of peak demand forecasting by training multiple learners and aggregating their predictions. The aggregation is achieved through strategies such as averaging, voting, and stacking [221], which exploit the complementary strengths of individual models. Ensemble methods can be broadly classified based on their generation process into sequential approaches and parallel approaches.

Boosting, a sequential ensemble method, focuses on iteratively correcting the errors of earlier models by giving greater weight to instances that were previously misclassified. This process continues until a predefined number of learners is generated or the learning criteria are met. Boosting algorithms, such as Adaptive Boosting (AdaBoost) and Gradient Boosting (GB), have proven particularly effective in addressing the complex dynamics of peak energy demand. Ahmad and Chen [222] highlighted the advantages of AdaBoost in forecasting load profiles across different time horizons, ranging from one month to one year. This study demonstrated that AdaBoost could effectively capture seasonal variations and dynamic load patterns, outperforming traditional ML models. Moreover, Zhang et al. [223] demonstrated the utility of GB in STPDF for Southern California. Their findings revealed the importance of the integration of renewable energy, identifying solar capacity as a key driver of peak demand. Further, Lu et al. [224] combined Extreme Gradient Boosting (XGBoost), an advanced variant of GB, with empirical mode decomposition techniques to predict daily peak load, achieving notable reductions in forecast errors. In practice, boosting frameworks typically rely on small learning rates for stability, with tree depth tuned to balance bias and variance. Subsampling of rows or features helps reduce correlation among learners, while early stopping is often applied to prevent overfitting. Although boosting excels at capturing complex interactions and handling imbalanced peak events, it can be sensitive to mislabeled outliers and distributional drift, requiring careful calibration and regularization [20,225].

Bagging, a parallel ensemble method, builds multiple learners on different bootstrap samples of the training dataset and combines their predictions to reduce variance and improve robustness. For regression tasks, bagging typically employs averaging or median aggregation to finalize predictions. Erick Meira and Fernando Luiz [226] utilized bagging to forecast the monthly load demand in countries at varying stages of development. Their innovative combination of bagging with exponential smoothing and SARIMA showcased the flexibility and effectiveness of the proposed methods in different contexts. Furthermore, their introduction of the remainder sieve bootstrap as a variant of bagging further improved forecast accuracy, yielding superior performance across diverse power systems. Bagging approaches can struggle when features are strongly correlated and sometimes underfit sharp extremes [148,150]. These limitations motivate refinements such as random forest, which extends bagging with random feature selection.

Random forest generates multiple bootstrap samples, constructs a decision tree for each, and uses voting for classification or averaging for regression tasks [221]. As the number of trees increases, random forest typically achieves lower generalization error and greater training efficiency compared to bagging. In practice, random forests are commonly configured with a few hundred trees, which provide a good balance between variance reduction and computational cost. To prevent overfitting, tree depth is usually kept shallow and minimum leaf sizes are enforced, while group-wise subsampling or principal component analysis can be used to decorrelate features and further stabilize predictions. Fan et al. [147] utilized random forest for peak demand forecasting, assigning weights via genetic algorithms and identifying random forest as a key contributor to improving the forecast accuracy. Furthermore, Wang et al. [227] applied random forest to predict hourly load patterns in educational buildings, outperforming regression trees and SVM, while highlighting the varying importance of features across different semesters. Furthermore, Berrisch et al. [9] developed an ensemble method integrating random forest with generalized additive models and deep neural networks, resulting in more robust peak demand forecasting models.

5.1.3. Probabilistic and explainable AI models

In the ML-driven peak demand forecasting stage, advancements in probabilistic and explainable AI models have significantly influenced peak demand forecasting by addressing the inherent uncertainties and improving interpretability in power systems. During the transitioning stage, early advancements in probabilistic methods laid the groundwork for contemporary approaches. Building upon these foundations, these methods have evolved from traditional probabilistic techniques to enhanced hybrid and AI-driven approaches, providing enhanced reliability and decision-making support.

The utilization of probabilistic forecasting became more popular in the 2010s and beyond, as computational resources and data availability improved. For instance, Atsawathawichok et al. [139] utilized Gaussian processes with multiple kernel designs to forecast long-term peak demand, optimizing feature-based kernels to enhance the forecasting accuracy. Moreover, Shabbir et al. [138] employed Bayesian additive regression trees and conditional kernel Gaussian process regression in conjunction with SVR to model contingency parameters, providing probabilistic forecasts that integrate socio-economic and environmental uncertainties. Compared with Gaussian processes, which encode smoothness and periodic structure via kernels and provide analytically calibrated intervals on small-to-medium datasets, and Bayesian additive regression trees that capture regime changes and interaction effects in mixed weather–calendar–economic inputs, we observe complementary strengths and failure modes. Gaussian processes may under-represent tail risk unless non-Gaussian likelihoods or non-stationary/heteroscedastic kernels are used, and their cubic scaling constrains large feeder-level deployments [138,139]. By contrast, Bayesian additive regression trees are robust to non-stationarity and categorical drivers but can be conservative near extremes and extrapolate poorly beyond the training support [138,201]. Across the reviewed studies, Gaussian processes tended to achieve lower distributional scores (e.g., CRPS) when relationships were smooth and covariates were dense and high-quality, whereas Bayesian additive regression trees performed better under structural breaks (policy or tariff shifts) and heterogeneous feeders dominated by interactions [62,139,201]. These complementary properties motivate combining kernelized baselines with tree-based Bayesian learners for stress-testing peak scenarios [138,201].

Quantile Regression Neural Networks (QRNN) have also gained prominence for their ability to estimate conditional quantiles, thereby offering a comprehensive probabilistic view of peak load variations. He et al. [64] demonstrated the efficacy of a hybrid model combining noise-assisted multivariate empirical mode decomposition with QRNN to produce probability density forecasts. This approach effectively decomposes high-frequency noise, leading to more accurate peak demand forecasts. cINNs represent another advancement in probabilistic forecasting. Heidrich et al. [125] introduced cINN for MTPDF, leveraging feature generation and statistical inputs to quantify uncertainty in both peak values and their timing. By providing a probability distribution instead of a single predicted value, this approach enhances risk management and informs capacity planning, offering a more comprehensive framework for decision-making.

In recent years, there has been a growing emphasis on not only improving performance but also ensuring the interpretability of advanced ML-driven models. Explainable AI techniques aim to demystify the decision-making processes of these models, enhancing trust and facilitating informed decision-making. Jang et al. [121] integrated deep neural networks with SHAP (SHapley Additive exPlanations) to explore feature importance in daily peak demand forecasting. This combination improved model transparency, offering stakeholders clearer insights into the factors influencing peak demand forecasting. Hybrid models that integrate probabilistic methods with explainable AI greatly benefit from the combined strengths of both approaches. Soman et al. [154] introduced a hybrid model integrating cINN with traditional neural networks to generate probabilistic forecasts while maintaining interpretability for peak demand forecasting. Moreover, Fu et al. [150] employed an ensemble ML approach combining random forest, gradient boosting machine, and logistic regression to predict the probabilities of peak day and peak hour. The proposed method incorporated data augmentation and feature engineering to provide probabilistic forecasts with clear explanations of influential factors. A practical gap remains between distributional forecasts and common explanation tools. Point-based attributions such as SHAP applied to expected predictions can hide how feature influence differs across quantiles, which is often where operators are most concerned with tail risk [121]. Explanations of probabilistic models also inherit Monte Carlo variability from posterior sampling, which raises questions about reproducibility, and the computational burden is considerable for Gaussian processes and Bayesian additive regression trees at the feeder scale. To address these challenges, the literature describes several practices. One is to report quantile-aware attributions so that feature effects are accompanied by credible intervals, and another is to pair global importance measures with calibration diagnostics such as PICP or CRPS, ensuring that explanations reflect the quality of uncertainty estimates [61,62,64]. Stability can be improved by fixing seeds and reporting the explanation variance across posterior draws. Explanations are also more relevant when they are aligned with operational loss functions, for instance, by weighting contributions according to the cost of missed peaks [66,150].

5.1.4. Advanced feature engineering and transformation

Advanced feature engineering has become increasingly important in ML-driven peak demand forecasting, leveraging a variety of techniques to extract, select, and transform relevant features from diverse data sources. These advanced methods improve the efficiency of forecasting models by effectively capturing and leveraging underlying patterns and dependencies within the data.

Mutual Information (MI) is a non-parametric method used to measure the dependency between variables, making it a powerful tool for feature selection in peak demand forecasting. For instance, Abera and Khedkar [228] employed CLARA (Clustering Large Applications) combined with MI to cluster appliance use patterns. This approach enabled the extraction of significant features related to appliance consumption and household demographics from smart meter data, thereby improving the forecasting accuracy. Moreover, Haq et al. [204] utilized MI for input selection in their hybrid model combining SVM, ANN, and K-Medoids. By identifying the most relevant features, the model could focus on the essential variables influencing peak demand, such as household demographics and appliance usage. Despite its utility, MI can be sensitive to dataset size and dimensionality, particularly in the presence of sparse or noisy data. To mitigate these limitations, studies in other domains have explored the use of conditional mutual information and kernel-based estimators that are better suited to small-sample scenarios. Additionally, variance-based and entropy-based filtering techniques have been employed as preprocessing steps to enhance the stability of MI computations [229,230]. Future peak demand forecasting models may benefit from incorporating these strategies to improve feature selection robustness under challenging data conditions.

Recursive Feature Elimination (RFE) is another advanced technique for feature selection that iteratively removes the least important features based on model performance. Yang [83] integrated RFE with prediction models to enhance the accuracy of next-day building energy consumption and peak demand forecasts. By systematically eliminating irrelevant or redundant features, the model retained only those variables that significantly contributed to the task, thereby reducing overfitting and improving generalization of the forecasting model. While RFE provides a robust framework for static feature selection, few studies have examined whether the features it removes are genuinely irrelevant. In addition, more advanced techniques are needed to address the complexities of time series and multi-variable datasets. One promising approach is to integrate RFE with L1 (Lasso) or L2 (Ridge) regularization within the underlying model to penalize overly complex solutions, helping ensure that retained features are both informative and generalizable [231].

DTW and Grey Relation Analysis (GRA) extend the concept of feature selection by focusing on specific characteristics of the data. DTW aligns time series data by measuring similarity between sequences, even if they are temporally misaligned, ensuring relevant patterns are identified for forecasting. Yu et al. [38] implemented DTW to analyze load pattern similarity before applying GRU-based models, ensuring that only the most relevant temporal features were used for forecasting daily peak loads. While DTW is ideal for handling temporal dependencies, additional techniques are required to manage diverse datasets with mixed variable types. For instance, DTW alone may struggle to capture multiple seasonalities and underlying trends in complex time series. To address these limitations, it can be integrated with seasonal-trend decomposition techniques or combined with multiscale decomposition to enhance temporal alignment and trend extraction prior to DTW-based matching [232]. GRA complements these methods by evaluating the correlation between variables and the target, ranking features to retain only those with the most significant influence on predictions. Guo et al. [202] incorporated GRA for variable selection in their hybrid model combining Fbprophet and adaptive Kalman filtering, effectively identifying the most influential weather and industrial production factors that affect monthly peak demand.

Moreover, advanced decomposition methods, such as empirical mode decomposition and its variations, have been extensively utilized to enhance feature extraction. Huang et al. [190] used multivariate empirical mode decomposition combined with particle swarm optimization and SVR to extract features from load time series, thereby improving the day-ahead peak demand forecasting. Further, Zhang et al. [21] applied ICEEMDAN (Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) in conjunction with LSTM to decompose the net load series, facilitating more accurate short-term net load forecasting by isolating intrinsic mode functions relevant to peak and valley predictions. However, mode mixing and sensitivity to noise remain significant challenges in empirical mode decomposition-based methods. Alternative approaches, such as variational mode decomposition or filter bank methods that offer greater control over frequency bandwidths and decomposition resolution, have remained largely unexplored in peak demand forecasting [233]. Future research should examine the potential of these methods to enhance feature extraction and improve forecasting performance.

Wavelet decomposition has been effectively applied in several ML-driven peak demand forecasting studies to capture both high and low-frequency patterns in load data. For example, Abdellah and Djamel [113] combined wavelet decomposition with ANN and genetic optimization to forecast daily peak demand. The decomposition allowed the model to reconstruct frequency components accurately, enhancing its ability to predict peak demand under varying conditions. Similarly, Panapakidis et al. [51] utilized the wavelet transform in combination with SVR to decompose the load series, enabling more precise day-ahead peak demand forecasting in the Greek power system. While wavelet decomposition is effective for extracting multi-scale features from non-stationary load time series, its performance is highly sensitive to the choice of wavelet basis and decomposition level. These choices can significantly influence the quality of the reconstruction and the consistency of the extracted features. Further research is needed to develop systematic criteria or adaptive methods for selecting wavelet parameters in the context of load forecasting.

These feature-engineering methods render large-scale and high-dimensional datasets more tractable. MI and RFE reduce wide covariate sets to the variables that carry the strongest signal. DTW aligns load histories so that comparable patterns remain visible even when timing diverges. GRA prioritises exogenous drivers, retaining only the most informative weather and industrial factors. Decomposition methods, including multivariate empirical mode decomposition, ICEEMDAN, and wavelets, disentangle trend, seasonal, and high-frequency components, allowing models to ingest denoised channels rather than raw series. To further enhance forecasting under non-linear and non-stationary conditions, multiple decomposers can be combined with deep embedding networks, enabling more adaptive representations tailored to the structural complexity of peak demand signals. In practice, these techniques reduce memory and compute load while preserving peak-relevant structures, thereby mitigating the curse of dimensionality and enabling models to process streaming data and support deployment in real time [138,207].

Table 12
Geographical application level at different forecast horizons.

	National	Regional	Municipal	Local	Total
STPDF	15	29	5	24	73
MTPDF	5	4	0	2	10
LTPDF	8	6	1	2	17
Multi	2	0	0	0	2
Total	30	38	6	28	102

5.2. Applications across different contexts

5.2.1. Geographical scalability

Geographical scalability defines the extent to which ML-driven peak demand forecasting methods can be adapted across different spatial levels. This subsection provides a detailed summary of the distribution of specified methods across various geographical application levels and forecast horizons based on the reviewed studies. The analysis maps the prevalence and application of peak demand forecasting methods within national, regional, municipal, and local contexts, as well as across short-term, medium-term, and long-term forecasting horizons. Table 12 summarizes the research distribution on geographical application levels across different forecast horizons.

At the macro level, national-level studies constitute the second most substantial segment. These studies address country-wide peak demand scenarios, integrating macroeconomic indicators, such as GDP, population, and industrial data alongside traditional load and weather variables. These exogenous variables are often included within regression-based, hybrid, or machine learning frameworks to capture their influence on peak demand. For example, Steinfeld et al. [80] employed regression-based methods to focus on the long-term peak demand characteristics of Sydney, Australia, providing policy recommendations for peak load reduction. Similarly, Atsawathawichok et al. [139] utilized Gaussian process models for LTPDF in Thailand, incorporating multiple kernel designs to improve prediction accuracy. The focus on national-level studies highlights the importance of peak demand forecasting in energy planning and policy formulation, where aggregated data and broader economic factors play a key role. However, while the importance of macro-economic factors is acknowledged, current national-level studies have rarely employed advanced econometric approaches such as vector autoregression or structural equation models. Instead, they remain grounded in time-series or ML paradigms. The complex interdependencies between macro-economic factors and peak demand have yet to be systematically modeled. Future research could therefore profitably integrate econometric methods with modern ML to improve causal inference and interpretability in national-level forecasting.

In contrast to the broader focus of the national-level studies, the regional-level studies, which focused on more specific contexts, constituted the largest segment, accounting for about 37 % of the total number of studies reviewed. These studies often focus on areas within countries, such as provinces, cities, or states, utilizing localized data to address unique regional load consumption patterns and infrastructural characteristics. For instance, Tairen et al. [201] conducted STPDF in Albuquerque, USA, and Huang et al. [190] employed the hybrid model for day-ahead peak demand forecasting in New South Wales and Victoria, Australia, highlighting regional variability in load patterns. The focus on regional levels allows for more detailed and context-specific forecasting models that can interpret regional economic activities, climatic variations, and infrastructure differences. While regional level studies provide valuable insights into local dynamics, they cannot ensure that the findings are representative of broader regional conditions. Datasets used in the studies are typically drawn from single utilities or service territories, with little attention to whether they reflect broader regional level conditions. Cross-geographic validation and pooled analyses remain rare, so findings from regional level studies often do not generalize beyond their original context, revealing a substantive gap. For rare extreme events, synthetic scenario generation and weather resampling can help expand sparse local records. When models are applied across areas with very different climates or demand structures, transfer learning or covariate shift adjustment may improve adaptability.

Zooming into more precise scales, there are fewer studies aimed at the municipal level, about 6 % of the total number of studies, focusing on citywide forecasts that incorporate municipal policy and tariff information. Nguyen and Manuel [137] demonstrated the application of Gaussian process models for STPDF in Austin, USA, integrating Bayesian updates for uncertainty quantification. This municipal focus is crucial for urban energy management, where local governance and infrastructure policies significantly influence peak demand patterns. The integration of municipal-level data enables the development of forecasting models that can support city-specific energy initiatives and sustainability goals. However current municipal level studies rely mainly on time series and machine learning models with weather and basic economic or tariff inputs. Only one study explicitly brings a policy lever into the input set through dynamic tariffs in Lisbon [118], while the potential impact of urban planning such as land use, density, and transport investments has not been taken into account. Incorporating urban planning and policy decisions as structured exogenous variables can strengthen municipal-level peak demand forecasting by linking electricity use more directly to land use patterns, infrastructure development, and policy interventions. In this context, two potential approaches are particularly promising. First, spatial econometric models, such as spatial lag, spatial error, and spatial Durbin, can capture spillover effects and quantify how local planning covariates propagate to citywide peaks. Second, integrating multi-agent models with forecasting frameworks allows heterogeneous household, commercial, and transport behaviors under different planning scenarios to be represented, producing more realistic demand trajectories and improving both the accuracy and policy relevance of municipal forecasts.

Moreover, local-level studies targeting granular scales, such as distribution feeders, microgrids, individual buildings, campuses, or specific facilities, account for approximately 27 % of the research. These studies often require high-resolution data to accurately reflect load variations and capture the complex operational dynamics of localized power systems. For instance, Haq et al. [204] employed ARIMA and K-Means clustering for STPDF on university campuses in Japan, effectively capturing and analyzing the unique energy consumption behaviors associated with educational institutions. Similarly, Waheed and Xu [207] integrate deep neural networks with feature selection for high-resolution peak demand forecasting in smart grid systems, highlighting the key role of localized data in improving the granularity of the forecasting. Given this reliance on high-resolution inputs, many local-level studies place particular emphasis on preprocessing to improve data quality. Common strategies include outlier detection and anomaly filtering (e.g., Isolation Forest in small-industry forecasts [20], DBSCAN (Density-Based Spatial Clustering of Applications with Noise) in substation studies [153]), clustering to smooth noisy load patterns or identify day types (e.g., K-means in building and campus level studies [85,195,197]), and dimensionality reduction to remove redundancy (e.g., PCA-based filtering in substation and feeder studies [105,192]). Some works further employ weather scenario generation [92] or demand-side management scenarios [42] to address data sparsity in extreme conditions. Local-level peak demand forecasting is critical for operational energy management, as it facilitates the implementation of precise demand-response strategies and supports the optimization of energy infrastructure at finer scales.

It is worth noting that a small subset of studies addresses multi-level geographical scalability, encompassing multiple spatial scales within a single forecasting model. These studies often employ complex methods to handle diverse data sources and spatial hierarchies. For example, Bovornkeeratiroj et al. [225] present a multi-method open-source toolkit integrating LSTM, SVR, and SARIMA for multi-horizon peak demand forecasting at both regional and national levels. Although multi-level studies highlight the potential for scalable forecasting models that can transition across geographical and temporal scopes, they remain significantly underexplored. Few studies have incorporated advanced techniques such as transfer learning or meta-learning to enhance model scalability and adaptability. Transfer learning allows knowledge to be transferred from data-rich regions to data-scarce ones, improving forecast accuracy in low-resource settings. Meta-learning goes further by enabling models to adapt rapidly to new locations or temporal resolutions through the identification of shared structures across related tasks. However, multi-level forecasting has yet to fully leverage these techniques in a consistent and scalable manner.

5.2.2. Forecasting horizon

The selection of appropriate forecasting methods is intrinsically linked to the forecasting horizon and the nature of exogenous variables, each playing a key role in accurately predicting peak energy consumption [8,92,145]. Fig. 9 illustrates the relationships between these variables and methods from the reviewed studies across STPDF, MTPDF, and LTPDF, highlighting the increasing complexity and diversity of variables as the forecasting horizon extends.

In the STPDF, weather-related variables emerge as the most influential factors across multiple modeling categories [73,143,155]. Notably, hybrid models exhibit a high association with meteorological inputs, such as temperature, humidity, or wind speed, demonstrating their capacity to capture volatile and dynamic weather patterns [51,55,189]. ANN models also show strong connections to weather variables, reflecting their effectiveness in learning non-linear relationships within short-term horizons [111,115,185]. Furthermore, calendar-related variables (e.g., day of week, holidays) hold substantial relevance at this time scale [134,195], particularly when using hybrid and neural network models that integrate cyclical trends and seasonal effects. Meanwhile, models like Bayesian networks or RNNs are sometimes employed for selective short-term applications, especially where probabilistic or sequential dependencies are crucial [133,153].

As the forecast horizon expands to MTPDF, the influence of exogenous variables increases. Weather remains an integral factor, but economic-related variables (e.g., GDP, industrial growth, demographic shifts) gain more significance, mirroring the increased complexity within this forecast horizon [157,184,192]. ANN and time series decomposition models are commonly used to capture fluctuations in climate conditions and cyclical demand trends in MTPDF [89,92,93]. RNN variants, including LSTMs, are also moderately correlated with medium-term weather and economic data, owing to their strength in modeling extended temporal dependencies [155,234]. Calendar and economic variables both play a substantial role in stochastic time series models [85,114], while Bayesian networks are recognized for capturing probabilistic dependencies among economic drivers that evolve over periods of weeks to months [138,206].

In the LTPDF, weather and economic-related variables become even more dominant. Hybrid models, which can integrate long-term climatic trends, demographic changes, and market or policy factors, are particularly effective in leveraging these inputs [8,139,140,164]. Regression-based models, ranging from simple econometric methods to non-linear or quantile regressions, also exhibit strong reliance on these variables, highlighting their proficiency in capturing linear and non-linear trends over the long-term [53,80,203,235]. Additionally, probabilistic or extreme value techniques are particularly relevant to long-duration planning scenarios, as they are adept at handling rare events and peak extremes [29,48]. Although policy-related variables are less used, they still play a meaningful role in regression and hybrid models, especially in cases where changes in policies such as renewable energy mandates or demand response programs create lasting structural impacts [42,66].

At present, only a very limited number of studies address multi-horizon peak demand forecasting, which aims to generate coherent

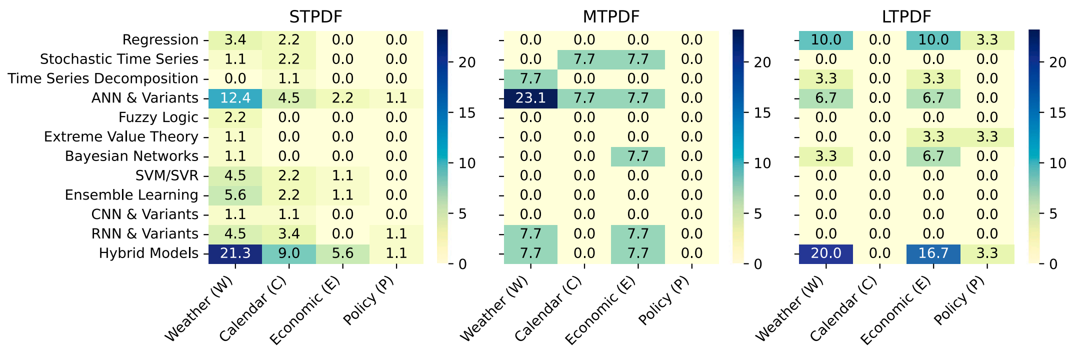

Fig. 9. Utilization of variables across forecasting methods and horizons.

predictions across different temporal scales such as daily, monthly, and annual peaks. For example, Jin et al. [206] proposed a hybrid LSTM-backpropagation neural network model that integrates multi-source information to jointly predict medium- and long-term peak loads in Guangdong, China. This approach leverages both neural sequence learning and feedforward structures to bridge time scales, while also incorporating weather, economic, and policy-related variables. Despite these initial contributions, multi-horizon forecasting remains substantially underexplored. Existing approaches often treat different horizons separately and lack mechanisms to maintain consistent prediction performance across time scales. Future research could advance this field by developing hierarchical frameworks that explicitly align forecasts across horizons, ensuring that aggregated short-term predictions converge with long-term planning outcomes. Multi-task learning represents another promising direction, where shared representations of variables such as weather, economic activity, and policy shifts can produce horizon-specific outputs while balancing short-term variability with long-term structural trends. In addition, probabilistic methods are well suited to represent how uncertainty propagates from short-term volatility to long-term structural change, which can also improve the ability of forecasting models to handle multi-horizon inputs. Incorporating these methods would support the development of robust multi-horizon forecasting systems capable of supporting unified planning and operational control.

In summary, in STPDF, the task is strongly shaped by immediate and volatile factors, such as weather and calendar variables, which are effectively addressed by hybrid, ANN, and DL models [111,196]. As the horizon extends to medium-term, economic considerations become more influential, prompting the adoption of methods such as stochastic time series and ANN for the MTPDF to track evolving socio-economic and seasonal trends [85,157,236]. For LTPDF, weather and economic variables become increasingly significant, with hybrid and regression-based approaches providing effective solutions for capturing long-term trends [8,66,139,164]. On the other hand, in LTPDF, deep learning models generally play a reduced role compared to medium or short-term contexts, largely due to sparse and coarse-grained input data, non-stationary patterns driven by structural changes, and a high risk of overfitting from limited training samples, highlighting the necessity of designing advanced deep learning frameworks capable of processing diverse inputs and modeling complex dependencies over long-term forecast horizons.

5.2.3. Stakeholders and scope correspondence

Effective peak demand forecasting supports multiple stakeholder groups, each characterized by distinct practical requirements and operational scopes. Aligning forecasting methods with these needs is essential for delivering actionable, relevant insights. To effectively capture these varied needs, forecasting models must be tailored in terms of feature selection, temporal resolution, and interpretability. For instance, residential, commercial, and industrial users differ not only in consumption patterns but also in operational priorities, necessitating stakeholder-specific configurations of inputs, model complexity, and performance metrics. This subsection delineates the main stakeholder categories and illustrates how various modeling approaches address their respective challenges.

Grid operators and electricity retailers typically manage national or regional forecasting tasks. Their concerns span grid stability, resource allocation, and infrastructure planning, necessitating high-accuracy, scalable methods that accommodate a broad range of exogenous variables. Ensemble learning, deep learning architectures, such as LSTM and GRU, and hybrid models frequently appear in this setting. Lee and Cho [8] describe a national-level approach using traditional machine-learning techniques, whereas Zhang et al. [21] discuss system-wide forecasting in Western Australia via hybrid frameworks. These models support both short- and long-term strategies by integrating meteorological, operational, and economic factors, enabling utilities to deploy informed solutions for peak load management.

Industrial and commercial entities operate on scales ranging from municipal to regional, often covering short- to long-term horizons. Their use cases include DSM, operational scheduling, and cost control, and they tend to adopt hybrid models, ensemble learning, and deep learning to handle real-time data and operational constraints. In the mining industry, Laayati et al. [149] introduce a fast forest quantile regression approach tailored for open-pit mine energy demand. Similarly, airport terminal energy systems benefit from a CNN-Transformer hybrid model that captures non-linear, dynamic load behavior [7]. Such setups facilitate effective forecasting and load optimization, reducing both energy costs and operational risks. Neglecting the operational priorities or constraints of these actors can lead to model outputs that are either inaccurate or impractical, resulting in costly inefficiencies and unexploited flexibility potential.

Residential buildings, characterized by localized consumption and high variability, require short-term, high-resolution forecasting solutions. Data often originate from smart meters and appliance-specific usage patterns, guiding the adoption of DL architectures (e.g., CNNs, LSTMs), fuzzy logic, and ensemble learning. An evolutionary LSTM ensemble in Ai et al. [22] focuses on household peak demand prediction, while Alduailij et al. [23] uses advanced neural networks to assess peak energy consumption in smart buildings. These refined techniques enable building managers and facility operators to optimize energy usage, reduce utility expenses, and maintain system stability. However, failing to incorporate residential perspectives, such as comfort preferences or behavioral variability, can hinder demand response program design and erode public engagement in flexibility initiatives.

Government agencies and policymakers require national, long-term forecasts that capture probabilistic and socio-economic variables. Their models often integrate Bayesian networks, Gaussian processes, or Explainable AI frameworks to capture uncertainty and provide transparent insights for policy decisions. Ploysuwan et al. [140] investigate peak demand forecasting for the electricity generating authority in Thailand, utilizing Gaussian processes at the national level. Meanwhile, studies by Lee and Cho [8] and Zhang et al. [21] emphasize the relevance of incorporating regional power system analytics into strategic policymaking. Such models supply probabilistic perspectives on long-term demand and accommodate shifts in economic and regulatory landscapes, reinforcing their value for energy policy formulation.

To support broader inclusion and ensure stakeholder trust, future forecasting frameworks should incorporate participatory mechanisms such as co-design workshops, survey-informed modeling priorities, or iterative validation with stakeholder feedback. These mechanisms ensure that critical modeling assumptions, such as risk tolerances, forecast granularity, or variable prioritization, are not imposed top-down but are informed by real-world constraints and priorities. Moreover, since stakeholder objectives may conflict, such as industrial reliability versus household affordability or utility-level stability versus distributed flexibility, forecasting models must be embedded within multi-objective planning frameworks. Techniques including hierarchical modeling, scenario analysis, or Pareto optimization can help balance competing demands, facilitating fair and efficient decision-making across system levels.

5.2.4. Economic trends in application locations

Economic trends play an essential role in shaping the adoption and implementation of peak demand forecasting methodologies in various geographical regions. The interplay between economic development, investment in energy infrastructure, technological advances, and policy frameworks significantly influences the focus and sophistication of the forecasting models used in different regions. Fig. 10 presents a global overview of where ML-driven peak demand forecasting methods have been applied.

At a broad level, the distribution of these forecasting initiatives indicates that advanced approaches are not limited to a single continent but are spread throughout the world. An overall trend is observed: countries investing heavily in energy infrastructure, data analytics, and the integration of renewable sources typically employ more advanced modeling methods. Their integration of cutting-edge ML technologies, which blend statistical and ML-driven techniques, alongside frameworks that align with policy highlights the significant impact of economic trends, regulatory incentives, and technological capabilities on determining methodological preferences. However, disentangling the causal relationships between these factors remains challenging. In many cases, energy infrastructure growth and forecasting capacity co-evolve with broader economic development. Future research could benefit from more sophisticated econometric frameworks that explicitly model the dynamic and non-linear interactions between economic indicators (e.g., GDP, electrification rate), infrastructure variables (e.g., grid expansion, smart meter penetration), and forecasting model adoption.

In countries with significant industrialization, such as the United States and Canada, increasing concerns regarding grid reliability and cost control are driving the advancement and implementation of enhanced forecasting systems [56,119,150,195]. These countries, supported by strong energy infrastructure, ample policy incentives for renewable energy integration, and well-funded research, frequently employ ML-driven and hybrid approaches, including deep neural networks, ensemble methods, and time series decomposition techniques, to achieve precise and scalable peak demand forecasting [60,121,154]. In this setting, standard methods often utilize feature extraction from weather, calendar, and operational data, as well as attention-based frameworks, to manage the complexity inherent in modern power systems. By utilizing cutting-edge models capable of addressing both short-term fluctuations and long-term patterns, system operators and policymakers enhance forecast robustness, flexibility, and interpretability, thereby facilitating more effective energy planning and grid management.

Meanwhile, in fast-developing countries, such as China, India, and Iran, the increasing energy demand driven by industrialization and urbanization encourages the use of advanced forecasting models that capture the interaction between economic expansion, changing consumption patterns, and renewable energy integration. Techniques such as hybrid models [184,192], and ensemble models [147,148] are increasingly adopted to tackle uncertainty. In India, for instance, models based on ANN [111] or non-parametric probability density forecasts [92] have proven effective in addressing the rapidly evolving nature of data and complexity. Recent multi-stage frameworks further refine mid-term electric load forecasting by integrating novel architectures in Madhya Pradesh [124]. These approaches enable more robust, flexible, and accurate forecasting under dynamic load behavior and diverse exogenous factors. However, in these contexts, the challenge of integrating disparate data sources, especially when data quality or coverage is uneven, remains significant. There is a growing need for adaptive models that can accommodate missing, noisy, or delayed inputs without compromising forecast reliability.

In emerging markets across other areas, the primary focus often lies on expanding electricity access, improving grid stability, and employing reliable forecasting techniques that can function effectively under such conditions. Local utilities are increasingly incorporating methods such as ANN and fuzzy logic, as well as econometric models tailored to smaller datasets and cost-sensitive environments. In Nigeria and Zimbabwe, for instance, annual peak demand forecasting leverages regression-based or econometric models [53,131], while countries like Algeria and Indonesia demonstrate the utility of adaptive hybrid methods and very

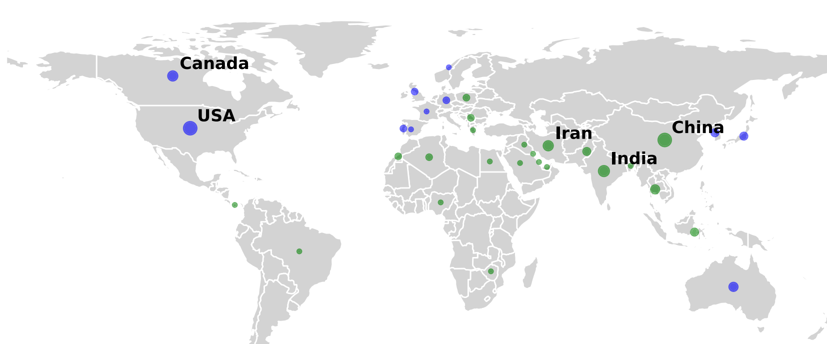

Fig. 10. Global projections of ML-driven peak demand forecasting. Each dot represents a country where ML-driven peak demand forecasting methods have been applied. Blue dots represent developed countries, while green dots represent developing countries. The size of each dot reflects the number of applications, with larger dots representing a higher frequency of applications. The top five countries with the greatest number of applications are the United States, China, Canada, India, and Iran.

short-term fuzzy logic to manage sudden load variations [93,134]. These regions often grapple with limited access to high-quality, granular data, which constrains the choice and performance of forecasting techniques. In such cases, robustness becomes paramount, requiring forecasting models that can tolerate imperfect or incomplete inputs through techniques such as imputation, ensemble averaging, and uncertainty-aware learning. Ignoring these data limitations can lead to systemic underestimation or overestimation of peak loads, potentially causing suboptimal planning decisions or resource misallocations. Therefore, the development of lightweight, noise-resilient models is essential for the equitable deployment of forecasting tools in low-resource settings. International collaborations and targeted funding further facilitate the adoption of these resource-efficient and scalable solutions, enabling operators to optimize demand-side resources, reduce power losses, and guide capacity planning in cost-effective ways [48,113,135,191].

In summary, the observed patterns reveal a robust interaction among economic growth, energy policy structures, technological investments, and the advancement of forecasting models. Regions investing in data collection systems, renewable integration, and digital infrastructure create conditions where advanced forecasting can effectively improve grid stability, inform infrastructure investments, and aid in meeting sustainability goals. Addressing regional disparities in data access and forecasting capacity is not only a technical issue but also a matter of energy equity. Future directions should consider participatory design processes that involve local stakeholders in the co-development of forecasting frameworks that reflect local realities, needs, and constraints. On the other hand, regions with more constrained resources often adopt phased approaches, gradually integrating more advanced and complex models to address emerging needs during their rapid development. This dynamic highlights the need to align advancements in peak demand forecasting methods with economic realities, resource availability, and policy goals, thus ensuring these methods remain effective and adaptable across various power systems.

5.2.5. Integration with renewable energy sources

The integration of renewable energy sources, particularly solar photovoltaics (PV) and wind turbines, has introduced both complexities and opportunities in peak demand forecasting. Unlike conventional generation, where output can be scheduled with relative certainty, renewables exhibit intermittent behavior due to fluctuating weather conditions and seasonal patterns. Even small deviations in solar irradiance or wind speed can cause significant swings in net load, thereby complicating traditional load prediction techniques. As a result, grid operators face increased pressure to balance supply and demand, avoid overloading, and maintain reliability while striving to achieve sustainability targets. These challenges underscore the need for robust and adaptive forecasting models capable of managing both rapid, short-term variability and long-term trends in renewable output [21,132].

To address these challenges, researchers and utilities have explored a range of advanced peak demand forecasting models tailored to renewable-rich environments. For instance, Faraji et al. [132] developed multi-year load growth models incorporating solar PV and wind turbines to capture temporal shifts in net demand and guide optimal planning of grid-connected microgrids in Tehran. In Western Australia, the EVT has been employed to assess how PV capacity affects long-term substation maximum demands, allowing utilities to identify high-demand scenarios under various solar adoption rates [29]. Some short-term net load forecasts combine LSTM networks with empirical mode decomposition to capture the granular fluctuations introduced by rooftop PV [21], while cINN has been proposed to model the complex interactions between load profiles, climate variables, and high-renewable adoption [125]. In parallel, studies focusing on DSM show that incorporating peak-shifting strategies and storage solutions (e.g., battery systems) can help mitigate the volatility introduced by intermittent renewables [42,66]. Together these works map onto several theoretical pathways for integrating renewable variability into peak demand forecasting. Some explicitly model demand-side behaviors through DSM and storage, while others focus on net demand by combining load and renewable generation forecasts. Probabilistic frameworks such as EVT and cINN quantify uncertainty, and the choice between forecasting in aggregated form or independently at the feeder level determines whether variability is smoothed or preserved. These dimensions highlight the different ways in which variability and uncertainty can be represented in forecasting frameworks under high renewable penetration.

Building on this, several approaches also extend the forecasting focus beyond single peak magnitudes. For instance, Zhang et al. [21] explicitly predicted both peak and valley values of net load, providing a measure of daily variability. DSM-oriented studies [42,66] demonstrate how load-shifting strategies reshape load duration curves, highlighting the importance of persistence alongside peak size. Similarly, the EVT analysis of substation demand under PV penetration [29] links peak estimation with return levels, effectively capturing the probability of sustained extremes. These studies show that outputs reflecting duration and variability are already emerging in renewable-rich contexts, and that they provide insights into flexibility and ramping needs which cannot be derived from peak values alone. The evidence collectively suggests that peak demand forecasting in renewable-rich systems increasingly requires attention to both the magnitude of peaks and their temporal characteristics, such as duration and variability.

Despite these advances, critical challenges remain. One difficulty is capturing the stochastic and weather-dependent nature of solar and wind generation, where short-term fluctuations can cause rapid swings in net load that even advanced models struggle to anticipate. Recent studies have begun to address this by incorporating extreme weather information into ensemble models [148] or by providing probabilistic forecasts that explicitly quantify uncertainty [63], but the problem of rapid intermittency remains unresolved.

Beyond methodological aspects, policy and economic drivers also play an important role in shaping how renewables interact with peak load. Incentives for DERs, such as rooftop PV, and the push for green energy have reshaped load curves, sometimes resulting in mid-afternoon net load valleys (‘duck curve’ as introduced by Azemena et al. [237]) or late-evening peaks. Accurate forecasting models must now factor in these shifting consumption and generation patterns, as well as the ramping capabilities needed to maintain grid stability. Forward-looking utilities and policymakers are incorporating demand response programs, real-time pricing, and capacity market mechanisms to encourage consumers to adjust their energy usage around renewable availability [42,66]. Methodological advances and policy efforts combined ensure that peak demand forecasting supports not only immediate operational efficiency but also long-term sustainability, resilience, and economic viability as renewables continue to expand across global power systems.

5.2.6. Advanced grid technologies

In modern power systems, advanced grid technologies, such as smart meters, Internet of Things (IoT) devices, and real-time communication networks, have redefined the way peak demand is measured, analyzed, and forecasted. These technologies offer high-frequency, granular energy-consumption data at the household, building, and distribution-feeder levels [204,228], revealing detailed load patterns that were previously obscured by the coarse resolution of traditional metering. By harnessing two-way communication between utilities and end-users, system operators can not only monitor near-real-time load conditions but also implement more responsive and targeted interventions. In this way, advanced grid technologies serve as the backbone of data-rich environments, where ML-driven peak demand forecasting models can be continuously refined to accommodate changing demand patterns and exogenous variables such as weather, holidays, or pricing signals [236].

A core advantage of these innovations lies in the capacity to develop dynamic, highly responsive forecasting models. With access to streaming consumption data, utilities can integrate ML algorithms to predict and respond to peak loads on a sub-hourly or minute-by-minute basis

[23,155]. In some cases, ensemble and hybrid methods combine time series decomposition, feature extraction, and real-time anomaly detection to capture abrupt shifts in load [20]. For instance, the rollout of IoT sensors in commercial and industrial facilities enables DSM programs that can shed or shift load in near-real time, mitigating spikes and enhancing grid reliability [20,42,66]. Moreover, at the utility or city scale, robust data flows facilitate geospatial load analyses and predictive maintenance, further reducing the risk of unexpected outages and volatile peak conditions [236].

However, the volume and velocity of data generated by advanced grid technologies create significant challenges for scalable management. An important issue is interoperability, since IoT devices and sensor networks often operate with heterogeneous protocols and data formats that require harmonisation through standardized interfaces and metadata schemes. Data quality is another concern, though many studies already mitigate missing values, noise, and misalignment through cleaning, filtering, and decomposition. In renewable and IoT-rich contexts, the high frequency and decentralized character of data streams make these problems more acute, and robust preprocessing becomes a prerequisite for integration into forecasting models. Emerging solutions include edge computing and hierarchical data architectures, which allow initial processing near the source to reduce bandwidth and latency, and federated learning, which enables decentralized training while preserving privacy. Alongside these techniques, advances in storage infrastructure and stream processing frameworks will be essential to ensure that forecasting remains accurate and timely at scale.

Looking ahead, IoT-based frameworks and smart grid innovations are poised to reshape peak demand forecasting models. By providing high-frequency, granular data from widespread sensor networks, these systems can reduce data latency and enhance the speed and accuracy of forecasting analyses. Integrating ML algorithms with real-time IoT data facilitates near-immediate situational awareness, improved grid reliability, and cost-effective load management. For instance, smart meter deployments enable dynamic, appliance-level monitoring [20, 228], while aggregated short-term forecasting techniques at the district level leverage high-resolution consumption data to refine peak forecasts [236]. As peak forecasting models become more reliant on massive, high-frequency datasets, investment in scalable stream-processing frameworks and intelligent data pipelines will be increasingly important. Techniques such as online learning and adaptive feature selection can further ensure that ML models remain both accurate and computationally feasible in real-time settings. These advanced digital architectures and adaptive forecasting methods offer utilities and policymakers unprecedented insights and control over grid operations, ultimately delivering more resilient and sustainable energy solutions [207].

5.3. Critical evaluation

5.3.1. Advantages of ML-driven peak demand forecasting

ML-driven methods have introduced several advancements that specifically address the unique characteristics of peak demand forecasting, which requires handling extreme variations and nonlinearities in peak demand data.

Modeling non-linear demand relationships. ML-driven peak demand forecasting models are particularly adept at capturing non-linear relationships between variables that traditional statistical models often fail to address. Recent frameworks developed to integrate multiple ML algorithms have further leveraged this ability to capture diverse nonlinearities. For example, Gajowniczek et al. [59] demonstrated how the integration of CART, K-NN, and ANN improved the classification of seasonal peaks using the strengths of each algorithm to process periodic variations in energy consumption. Similarly, Liu and Brown [56] proposed a multi-algorithm framework combining Naïve Bayes, SVM, random forest, AdaBoost, CNN, and LSTM, enabling the model to isolate intricate demand patterns and provide accurate forecasts under diverse scenarios. Furthermore, open source tools like PeakTK by Bovornkeeratiroj et al. [225] have introduced platforms that integrate diverse forecasting methods and datasets into a single framework. By providing a standardized tool for evaluating and benchmarking forecasting methods, such platforms facilitate advancements in both model development and application.

Adaptability to diverse variables. Peak demand forecasting must consider highly dynamic and context-specific drivers, such as extreme weather events and large-scale public gatherings. These transient events create abrupt changes in energy consumption, necessitating models that can dynamically adjust to such variations. Advanced feature extraction techniques, including RFE and PCA, play key roles by isolating the most important variables while reducing noise, ensuring models remain focused on impactful predictors [26,148]. Besides, integrating multiple algorithms has proven effective for adapting to heterogeneous energy consumption patterns. For example, Bellahsen and Dagdougui [236] demonstrated that combining LSTM, random forest, and K-NN enabled accurate building-level peak demand forecasting by leveraging the strengths of each approach. LSTM networks are good at modeling sequential dependencies in time series data, while random forest handles categorical and numerical data robustly, and K-NN captures local variations. This combination facilitates better performance when forecasting demand across varied building types and usage patterns.

Handling extreme events and probabilities. Rare yet impactful events, such as heatwaves and industrial surges, pose significant challenges for peak demand forecasting due to their deviation from typical demand patterns. Probabilistic approaches like EVT and Bayesian networks are well-suited for modeling such outliers. EVT focuses on the statistical properties of extreme demand events, enabling accurate predictions of peak probabilities during rare occurrences [46]. Bayesian networks, on the other hand, model probabilistic relationships between variables, providing insights into the dependencies that drive peak loads [138]. Incorporating these methods into broader ML frameworks enhances their predictive capability, particularly for long-term forecasts where extreme events can have a disproportionate impact. For example, combining EVT with ML-driven models improves reliability in scenarios involving high-impact demand fluctuations [235]. In addition to EVT and Bayesian networks, ensemble methods can also serve as practical tools for uncertainty quantification. By aggregating predictions across diverse learners, ensembles provide distributions of outcomes rather than point estimates, allowing decision-makers to evaluate the confidence intervals of peak forecasts. This complements Bayesian approaches by addressing both epistemic and aleatoric uncertainty in operational contexts.

Hybrid modeling for enhanced seasonal dynamics. Hybrid modeling provides flexibility in addressing diverse demand patterns across different contexts. It is worth noting that, based on the reviewed studies, integrating ML with traditional decomposition techniques has proven highly effective in addressing the complexities of seasonal dynamics in peak demand forecasting. For instance, the combination of wavelet analysis with ANNs allows for the decomposition of data into different frequency components [196]. This approach improves the ability of the model to capture seasonal shifts and long-term trends, resulting in more precise peak demand forecasts. Moreover, Amara-Ouali et al. [238] explored multi-resolution modeling, which integrates generalized additive models and ANN, highlighting how decomposing time series data isolates periodic drivers of peak variations, enabling models to adapt dynamically to changes in consumption patterns. These decomposition–ML hybrids function as adaptive smoothing schemes, mitigating the limitations of fixed-coefficient exponential smoothing by learning time-varying seasonal structures and non-linear cross-effects.

Dynamic forecasting with temporal models. Dynamic forecasting requires models capable of capturing the temporal dependencies inherent in energy consumption patterns. Deep learning models such as RNNs and LSTMs have proven particularly effective in this domain. For instance, Yu et al. [38] demonstrated the effectiveness of RNNs in STPDF, achieving high precision by modeling temporal variations and incorporating recent load data. Similarly, Ibrahim and Rabelo [234] employed deep architectures, including CNN, CNN-LSTM, and multihead CNN, for MTPDF. These models captured complex behaviors in load profiles across multiple timescales, further emphasizing the importance of temporal adaptability in handling dynamic demand scenarios.

Integration of real-time data. ML-driven methods are highly effective in integrating continuous real-time data streams from smart grids, enabling more adaptive and responsive peak demand forecasting. Real-time inputs, such as renewable energy generation levels, battery storage capacity, and consumption data, allow forecasting models to adjust dynamically to changing grid conditions, supporting timely decision-making for peak load management. Through advanced feature engineering and transformation, real-time data is reduced to a compact set of denoised signals, enabling efficient, low-latency forecasting. For example, Shabbir et al. [138] highlighted the improved accuracy of probabilistic models when incorporating real-time data from smart grids, showcasing the value of dynamic updates in managing fluctuating demand scenarios. Similarly, Waheed and Xu [207] demonstrated how deep learning models can learn incrementally from real-time energy consumption data, improving the adaptability of forecasts to sudden shifts in demand patterns.

5.3.2. Limitations and challenges

Despite these advantages, ML methods face specific challenges that arise from the nature of peak demand forecasting.

Complexity of short-term extreme variations. While ML-driven peak demand forecasting frameworks demonstrate significant advancements in capturing non-linear relationships and diverse patterns, short-term extreme variations, such as abrupt demand surges, remain challenging due to their unpredictable nature and high-frequency dynamics. Traditional ML frameworks, even those enhanced with deep learning or hybrid approaches, may lack the responsiveness required for such high-frequency changes. Models designed for short-term horizons, such as CNN-LSTM architectures, demonstrate strengths in capturing temporal patterns but encounter difficulties when applied to scenarios with unexpected demand spikes, as highlighted in studies on dynamic peak load behaviors [56,234]. The inherent complexity of these variations requires not only advanced temporal modeling, but also a robust mechanism for incorporating real-time data to adapt to rapid shifts.

Overfitting in peak scenarios. Overfitting poses a major challenge in ML-driven peak demand forecasting and can be broadly observed in two forms. The first type arises from data sparsity, particularly for rare and extreme peak events. In such cases, models overly rely on limited historical patterns, which restrict their ability to generalise to novel conditions. For instance, Deng et al. [148] noted that boosting-based learners required careful regularisation to avoid excessive focus on specific past events, while Hsu and Chen [50] found that ANNs, although effective at capturing non-linear relationships, struggled to generalise under sparse peak data. Hybrid approaches such as wavelet–ANN models also risk overfitting by overemphasising decomposed seasonal features [196]. Mitigation strategies here require strategies like regularization (e.g., L1/L2 or dropout), robust cross-validation, and data augmentation through synthetic event generation [148]. These methods ensure better generalization, enhancing the reliability of ML models for dynamic and evolving power systems. In addition, addressing overfitting in peak demand forecasting requires careful control of input dimensionality, since high-dimensional data can worsen the risk of memorising noise. Techniques such as Bayesian regularisation [110], recursive feature elimination [83], and cascaded or hierarchical neural networks [70,180] reduce the effective dimensionality of the input space and improve generalisation, thereby preserving transparency in the modeling process. The second type of overfitting is design-related and emerges in structurally complex hybrids. When multiple components are combined, such as ANN fuzzy hybrids or deep learning-based hybrids, the increased capacity and large number of hyperparameters can lead to memorising noise rather than capturing generalisable patterns. Studies have addressed this through evolutionary optimization (e.g., GA or hybrid modeling for enhanced seasonal dynamics) [181,190], careful parameter and hyperparameter tuning [47,200], and embedding physical or economic constraints as regularisers [42,202]. These two forms of overfitting show that model design must be adapted to the characteristics of the available data and to the length of the forecast horizon. Operational constraints also influence the choice of modeling strategy. Because forecasting settings differ widely, no single architecture or training procedure performs well across all conditions. In peak demand forecasting, effective model selection therefore relies on benchmarking across representative use cases, on testing the sensitivity of results to model complexity, and on ensuring that outputs align with the needs of downstream decision-making. Recent studies increasingly use modular hybrid designs that balance predictive accuracy with interpretability and computational cost, pointing to the value of approaches that are tailored to specific contexts rather than generic solutions.

Data gaps and rare event sampling. A recurring limitation in peak demand forecasting is the incomplete or inconsistent availability of high-quality data, which can be considered in three main folds. First, the granularity of load data can frequently present issues. Resolution and aggregation levels vary across studies, from hourly system-level data to fine-grained smart meter or feeder measurements, and geographic coverage is not always consistent. Models trained on coarse or aggregated data struggle to capture localized peaks or short-term fluctuations [68,201]. Addressing this limitation requires improved access to high-resolution load measurements [199,207], hierarchical forecasting methods, and transfer learning techniques to bridge gaps across regions. Second, many models rely heavily on external drivers such as precise weather observations, calendar effects, and detailed economic indicators. However, these exogenous variables are not always recorded uniformly or at the same resolution [122,145]. Missing weather data, unlogged calendar events, or incomplete socioeconomic records can lead to gaps in the training dataset, reducing model accuracy and reliability. The reliance on detailed external data is therefore a common shortcoming in peak demand forecasting studies, since many utilities cannot guarantee uniform availability. Solutions explored in the literature include robust imputation, multi-source data fusion, and probabilistic treatment of missing inputs [20,118]. Third, the rarity of extreme peak events presents a major challenge, such as spikes caused by severe heatwaves or sudden economic shifts. These events may only occur a few times each year (or less), limiting the number of samples available to the model. ML-driven peak demand forecasting models can struggle with such imbalanced data because they tend to bias predictions toward more frequently observed conditions [21,42,48]. As a result, capturing the nuances of rare peak-demand extremes requires specialized approaches, such as advanced oversampling, synthetic data generation, or the aforementioned EVT, to ensure that the forecasting model accounts for the small-but-important fraction of cases that drive system stress and resource planning decisions.

Extreme-event uncertainty quantification. Techniques for extreme event modeling, such as EVT and Bayesian networks, are indispensable for quantifying risks associated with rare but impactful peak load events. However, these methods are computationally intensive, mainly due to their reliance on parameter estimation, probability calculations, and simulation-based validation. Belzer and Kellogg [46] emphasized the computational burden introduced by implementing EVT for daily peak demand modeling, where Monte Carlo simulations were essential to handle the probabilistic nature of extreme demand events. While these simulations improve predictive reliability, they significantly increase processing times and resource requirements, making them challenging to integrate into real-time applications. Similarly, ensemble methods often combine multiple models to improve accuracy, but can compound computational costs during both the training and inference phases [225,236]. These approaches require processing high-dimensional data, frequent updates, and iterative optimization, further exacerbating the resource demands. Beyond computational aspects, extreme-event modeling raises several practical concerns. First, EVT performance is highly sensitive to assumptions of independence and threshold choice. Multi-day heatwaves and demand-side interventions often induce dependence and non-stationarity, which makes declustering, threshold diagnostics, and covariate-driven non-stationary EVT advisable [29,42,48]. Second, when assessing coincident peaks, Bayesian networks enable data-efficient, policy-aware 'what-if' analysis but can understate joint tail co-movement. Copulas, including vine copulas, capture tail dependence more effectively but require larger datasets and are harder to interpret. Third, density forecasts also require systematic auditing with measures such as the aforementioned PICP and CRPS. They are best communicated through peak-oriented quantiles in the 90–99 % range, combined with blocked or seasonal cross-validation to address the class imbalance problem in peak demand forecasting [62,64]. Finally, in climate-resilient planning, scenario-conditioned posteriors that incorporate temperature, electrification trends, tariff structures, and DSM variables allow uncertainty estimates to be aligned more closely with operational decisions [8,66,139,140].

Integration with decentralized power systems. As the penetration of DER increases, such as rooftop PV panels, wind turbines, and home battery storage, grid operators and utilities face new challenges in accurately forecasting peak demand [22,132]. Traditional peak demand forecasting methods often assume centralized generation of load profiles, but decentralized systems introduce more variability and uncertainty. For example, a sudden reduction in solar power due to cloud cover or a coordinated discharge of battery storage systems can shift or flatten load peaks in ways not captured by conventional models. Moreover, the time-dependent interaction between variable renewable generation, storage units, and consumption patterns requires models that can dynamically adapt to network constraints and localized microgrid schedules [21]. As a result, effective forecasting in such environments calls for enhanced approaches that capture localized consumption patterns, integrate intermittent renewable production, and incorporate operational constraints such as battery charging schedules or demand response signals. Another related limitation in decentralized environments is the heterogeneity of data. Smart meters, rooftop PV inverters, IoT sensors, and weather stations provide information at different temporal resolutions and with varying reliability. Reconciling these heterogeneous inputs remains challenging and calls for advanced preprocessing, multi-source fusion, and hierarchical modeling strategies.

Interpretability and stakeholder trust. Amid the rise of more complex deep learning and ensemble algorithms for peak demand forecasting, questions of interpretability have become important [74]. Utilities, policymakers, and end-users must comprehend not only the forecasted load values but also the underlying drivers influencing these predictions, such as weather fluctuations, calendar effects, economic trends, or changes in consumer behavior. In black-box models where internal decision-making processes are opaque, it becomes challenging to justify the forecast to regulators or to adapt operational strategies based on model insights. This lack of explainability can undermine stakeholder confidence and slow the adoption of cutting-edge forecasting technologies. Decision-oriented explanations can help address this gap by clarifying which variables contribute to the predicted peak and how sensitive the forecast is to changes in those inputs. For utilities, this information supports procurement and demand-response scheduling. For policymakers, it offers a basis for tariff setting and targeted interventions. For consumers, it makes peak drivers more transparent and encourages participation in demand-side programs. Recent work has begun to introduce explainable AI techniques and model-agnostic methods such as SHAP, partial dependence plots, and rule-based decision visualization [121]. These approaches shed light on the contribution of different input variables, highlight periods of heightened risk, and allow operators to trace how and why the model arrives at specific peak forecasts. It should be noted that interpretability is necessary but not sufficient for stakeholder trust. Forecasts also depend on reliable data, regulatory alignment, and workable system integration. Where data are inconsistent, for example, with gaps in smart meter coverage or variable sensor resolution [122,145], even transparent models may produce insights that are not actionable. Regulatory frameworks likewise expect decisions to be justified, which many ML models cannot yet provide without additional explanatory layers [53,121]. Integration with SCADA (Supervisory Control and Data Acquisition), energy management systems, and tariff-setting systems presents further challenges, since these infrastructures were not designed to handle outputs from complex or probabilistic learning models [42,236]. A further issue is the correlation among input variables. Weather, calendar, and economic factors often overlap, creating multicollinearity that weakens model reliability. Techniques such as PCA, regularization, and ensemble learning help reduce this risk by limiting redundancy. In developing regions, the scarcity of exogenous data makes the problem more acute, and researchers have turned to transfer learning and data augmentation to compensate for missing information [239,240]. Yet even with these approaches, models remain constrained by theoretical limits, since capturing complex non-linearities requires representative data and sufficient sample size, without which generalization is difficult to achieve.

6. Future directions

The energy sector today is influenced by the swift integration of new technologies, such as blockchain-enabled peer-to-peer trading, IoT-driven smart metering, and advanced deep learning architectures, which introduce new forms of demand variability and high-resolution data streams. At the same time, the rise of edge computing, federated learning, and multi-agent systems presents new opportunities to decentralize and scale forecasting frameworks while respecting privacy and system heterogeneity. However, these advancements also introduce pressing challenges, including the need for explainable models, geographic generalizability, and seamless integration with DERs. As a result, the future of peak demand forecasting lies in developing adaptive, interpretable, and resilient systems capable of learning from diverse, real-time data while supporting decision-making across increasingly complex grid environments. This section highlights existing research gaps that require deeper investigation to address current limitations, as well as emerging trends that are shaping the evolution of the field.

6.1. Research gaps

6.1.1. Data-efficient generalisation

A persistent challenge in peak demand forecasting is how to build accurate models under conditions of scarce, noisy, or fragmented data. This issue is particularly acute in newly developed urban areas, microgrids, or regions with limited metering infrastructure, where historical records are insufficient to support complex machine learning models.

Transfer learning provides one potential pathway to address these limitations. It refers to the practice of applying models or learned representations from one domain or dataset to another, thereby reducing the need for large amounts of target data. The feasibility of this approach lies in the observation that peak demand patterns are often shaped by common drivers, such as temperature, calendar effects, and daily activity rhythms, which tend to repeat in different contexts. For instance, Cai et al. [239] demonstrated that transfer learning can improve short-term load forecasting in data-scarce regions by transferring deep neural network features learned from data-rich regions. Similarly, Li et al. [240] proposed a multi-source transfer learning ensemble LSTM method that selects similar source buildings and integrates their knowledge to improve multi-load forecasting accuracy for target buildings with limited data. These studies highlight the potential of transfer learning to reduce data requirements while maintaining accuracy, making advanced ML methods more accessible to regions with limited resources.

At the same time, geographic diversity poses additional challenges for generalisation. Most peak demand forecasting studies have been conducted in developed regions such as North America, Europe, and Australia, which benefit from long and reliable datasets, well-established metering infrastructure, and relatively stable grid operations. By contrast, developing regions in Asia and Africa often face data scarcity, inconsistent reporting standards, and unique grid infrastructures that limit the application of advanced forecasting techniques [48,53,131,191]. In these contexts, deep learning and hybrid models are less frequently applied due to insufficient historical data and the absence of essential explanatory variables such as comprehensive economic indicators or real-time sensor measurements [66,135,235]. Moreover, variations in policy mandates and regulatory frameworks across countries can hinder systematic data collection and open data sharing, posing challenges to the broader application of complex forecasting methods [73,234].

Addressing these constraints requires a coordinated strategy across modeling, data governance, and deployment. On the modeling side, data-efficient yet robust approaches such as hierarchical learning or transfer learning that borrow strength across comparable regions are needed, complemented by probabilistic formulations that capture uncertainty in short and noisy demand histories. On the data side, initiatives such as anonymised dataset sharing, the establishment of uniform quality standards, and training for local practitioners can improve coverage, comparability, and trust in results [113,125]. Finally, deployment frameworks must reflect regional constraints by adopting privacy-preserving collaborative learning and lightweight edge implementations that respect limitations in bandwidth, computation, and data-sharing norms [235].

6.1.2. Context-aware and explainable forecasting

A central challenge in peak demand forecasting is not only to predict extreme loads accurately, but also to ensure that forecasts are transparent, contextually grounded, and usable for operational decisions.

Incorporating structured domain knowledge into machine learning models offers substantial benefits, especially in capturing context-specific anomalies and system behaviors that are difficult to learn from data alone. For example, planned maintenance, large public events, or demand-response activations can disrupt regular load patterns, yet these are rarely encoded in historical datasets. Miller et al. [241], for instance, combined macroeconomic trends and temperature-related factors within a fuzzy logic based system to improve long-term forecasts. Formalizing such knowledge into structured inputs is, however, time-consuming and requires cross-disciplinary collaboration, while models must be carefully designed to balance domain-driven and data-driven signals.

Beyond structured knowledge, multi-source data streams further strengthen forecasting. Natural language processing can extract semantic features from policy announcements, news, and social media, which signal demand shifts and have been shown to improve day-ahead forecasts [242,243]. Likewise, computer vision can capture structural drivers of peaks such as building density, rooftop solar, or EV charging infrastructure [244,245], though its application to peak forecasting remains unexplored. In related energy studies, these textual and visual signals have been used to trace policy changes, monitor infrastructure growth, and detect sudden demand shocks.

Explainability ensures that enriched inputs translate into actionable outputs. Operators must understand the factors elevating exceedance risk, the reasons for intervention activation, and the methods to diagnose unforeseen peaks. Recent advancements, such as the RAID framework by Jang et al. [121], showcase the use of SHAP values to improve the transparency of deep learning models by highlighting key variables such as past peak loads and temperature. However, achieving an optimal balance between performance and transparency remains a challenge. While RAID improves interpretability, it still relies on complex neural networks that can be difficult for stakeholders to fully understand. Beyond RAID, other hybrid models have shown promise [200]. These approaches aim to clarify relationships between peak demand drivers, such as weather patterns and historical load data. However, their application often lacks generalization, making them less practical for diverse energy forecasting scenarios. Several promising methods remain underexplored in current peak demand forecasting practice. For instance, local interpretable model-agnostic explanations can isolate the specific drivers behind individual forecasts, providing localized insight into model behavior. Similarly, neural additive models learn per feature response curves that can be audited or constrained to reflect plausible demand–driver relationships [246,247]. By explicitly revealing how predicted peaks vary with controllable inputs, such intrinsically interpretable models offer more actionable insights than post-hoc attribution alone. To address these challenges, future research should focus on developing explainable frameworks that deliver consistent and actionable insights without sacrificing predictive accuracy, to ensure broader scalability and stakeholder trust. Further, future work should also focus on integrated pipelines where domain knowledge, textual and visual signals are combined with explainable frameworks, producing forecasts that are both robust to anomalies and transparent to stakeholders.

6.1.3. System-level architectures

System-level architectures for peak demand forecasting increasingly rely on decentralized and distributed intelligence, where agents, edge devices, and blockchain infrastructures collaborate to process data and generate forecasts in real time. These approaches move beyond centralized pipelines and better reflect the heterogeneity and autonomy of modern power systems.

Multi-agent systems refer to decentralized architectures composed of multiple autonomous agents that interact, collaborate, or compete to achieve individual or collective goals. Multi-agent systems can accommodate heterogeneous data streams from smart meters, IoT devices, and distributed resources, turning heterogeneity in peak demand forecasting into a modeling feature rather than a limitation. Traditional forecasting models, while effective for pattern recognition, often fail to capture the emergent spikes driven by synchronized behaviors across diverse consumers. Multi-agent systems, by contrast, can represent households, appliances, and distributed resources with decision-making capabilities and stochastic variability. For example, Ali et al. [248] demonstrate how EV battery managers and load agents interact within a multi-agent system to optimize dispatch and contribute to peak shaving, while peer-to-peer demand response systems coordinate flexible loads without centralized control [249]. These examples show how multi-agent systems can simulate localized load behaviors and interactions that give rise to peaks, offering insight into the mechanisms behind demand surges and enhancing interpretability. Moreover, multi-agent system frameworks support scenario analysis of tariff changes, EV penetration, or demand response programs, and can generate synthetic data or integrate adaptive learning agents to bridge behavioral realism with predictive precision.

Edge computing enhances real-time processing by bringing data processing closer to the source, thereby reducing latency and improving security. Promising techniques include clustering-based preprocessing methods such as K-Medoids on edge devices to reduce data dimensionality [204], and neural networks adapted to constrained environments for high-precision load prediction [207]. However, several limitations hinder large-scale adoption. The diversity and volume of data across regions challenge integration, while device constraints limit model complexity and create trade-offs between accuracy and responsiveness [164,199]. A further difficulty is the lack of standardized protocols, which complicates interoperability across infrastructures [206,234]. Recent work suggests that distributed databases and stream-processing frameworks can complement edge nodes by enabling scalable ingestion and synchronization of high-frequency, heterogeneous data streams [225,238]. In practice, edge devices are best suited for local filtering and dimensionality reduction, while distributed systems maintain consistency at larger scales. Promising directions include model compression (pruning, quantization, knowledge distillation) and adaptive online learning, which allow models to remain efficient, responsive, and privacy-preserving in edge environments.

Decentralized peer-to-peer energy trading introduces volatile consumption patterns and new load variables, significantly expanding the complexity of forecasting frameworks. While integrating real-time transactions, consensus mechanisms, and distributed ledger data into peak load models offers substantial theoretical and practical opportunities, it has not yet been comprehensively addressed in current methods. Existing models for peak demand forecasting often overlook blockchain-driven demand variability. Traditional regression-based methods, such as those applied to regional systems, focus on static historical variables and cannot capture decentralized exchanges [80]. Hybrid deep learning models that integrate feature transformation provide more accurate forecasts under such conditions by clustering consumption patterns and dynamically adapting to blockchain transactions [73,185]. Future research may combine blockchain with edge computing, where local aggregation of transaction data reduces latency before synchronizing with distributed ledgers. Embedding forecasting models directly within smart contracts or consensus mechanisms could further improve autonomy, allowing forecasts to update dynamically as new transactions occur. To ensure scalability and trust, lightweight privacy-preserving methods and federated learning can be incorporated, while on-chain recording of forecast outputs enhances auditability, reproducibility, and transparency. By embedding interpretability and traceability into blockchain workflows, decentralized systems can support more accountable and participatory decision-making among prosumers, utilities, and regulators.

In summary, multi-agent systems, edge computing, and blockchain-based trading can be viewed as complementary strategies for system-level forecasting. When combined, they support forecasting frameworks that are decentralized in structure, adaptive to changing operating conditions, and transparent in their market interactions. Such designs link predictive capability directly to the operational requirements of grid management and to the participatory role of end users in future power systems.

6.1.4. Integration with advanced DERs

While certain studies examine the integration of renewable energy and storage [21,238], a comprehensive analysis of the modeling of electric vehicles, distributed generation, and battery systems remains necessary. These flexible resources serve as dynamic elements that increase the complexity of the forecast by presenting a variety of diverse and variable consumption behaviors [22,205]. Furthermore, modern DERs require the incorporation of various data sources, such as smart meter readings, renewable generation outputs, and electric vehicle charging patterns, adding complexity to data preprocessing and feature engineering [150,225]. Studies such as Waheed and Xu [207] and Heidrich et al. [125] highlight the difficulties in handling and integrating data from solar PV systems and battery storage, which are essential for precise peak demand forecasting.

Moreover, deploying DERs across various application levels, from residential to industrial, requires scalable forecasting models that can accommodate diverse operational environments and ensure data consistency across distributed nodes [20,125]. The scalability challenge is highlighted in studies across different geographic and infrastructure contexts, such as smart grid systems [207] and industrial implementations [149]. Future research will therefore need to focus on scalable frameworks that combine data fusion with algorithm optimization, enabling EVs, distributed PV, and storage to be modeled in an integrated way and their impact on peak demand to be more reliably captured.

6.2. Emerging trends

6.2.1. Advanced deep learning techniques

Transformer-based architectures and attention mechanisms have emerged as powerful tools for capturing long-range dependencies and complex temporal patterns in peak demand forecasting. Hybrid models, such as CNN-Transformer combinations [7] and dual-attention LSTM encoders [156], demonstrate enhanced capabilities in long-term and multi-scale demand predictions. Additionally, sequence-to-sequence models [155], and hybrid CNN-RNN architectures [164] are increasingly adopted to extract both spatial and temporal features, thereby improving prediction accuracy for peak demand.

Another notable trend is the integration of probabilistic forecasting methods with deep learning models [63] to address uncertainties and provide more robust peak load predictions, utilizing models such as cINN [125] to generate probability density functions of peak loads. Furthermore, advancements in lightweight deep learning architectures, including model pruning and efficient neural network variants such as GRUs [38], are important for deploying real-time forecasting models on resource-constrained edge computing infrastructures. By optimizing algorithms for scalability and efficiency, these techniques enable the seamless integration of heterogeneous data from DERs and smart meters, thereby enhancing the resilience and adaptability of peak demand forecasting systems [20,235].

6.2.2. Federated and distributed learning

With the increasing availability of high-resolution load data, such as residential smart meter data, data privacy has become a major concern in peak demand forecasting. Protecting consumer data using private encryption algorithms is an essential task for researchers in the digital era [250]. The challenges related to electricity data transmission, storage compliance, security, and privacy protection continue to impede effective data utilization [251]. Additionally, the size and quality of the data significantly influence the training quality of ML models. However, relevant data necessary for accurate forecasting are often distributed across different organizations, each holding data that is subject to privacy laws restricting full data sharing. Aggregating all data into a centralized third-party database could facilitate the construction of robust forecast models, but it introduces significant security risks due to the centralization of sensitive information [252].

Privacy-preserving strategies, such as federated learning, offer a solution by allowing model training on decentralized data without requiring central storage of raw data [151]. This approach is particularly beneficial for grid operators seeking collaboration across multiple utilities or regional bodies while adhering to data privacy, security, and regulatory requirements. By leveraging federated learning, diverse data sources from different organizations can be combined to build accurate forecast models without compromising data privacy. This is achieved by training local models on-site and only sharing model updates (rather than raw data) with a central server, allowing collaborative learning across utilities while preserving confidentiality, ensuring compliance with data regulations, and maintaining scalability through parallelized training. Therefore, designing forecast frameworks that incorporate privacy-preserving techniques such as federated learning represents a promising future research direction to improve peak demand forecasting [253–255].

6.2.3. Reinforcement learning and adaptive systems

Dynamic decision-making through Reinforcement Learning (RL) presents a promising avenue for peak demand forecasting by enabling models to make real-time adjustments based on evolving grid conditions. Although still in the nascent stages, RL-based approaches can potentially enhance the adaptability and responsiveness of forecasting systems. Studies integrating evolutionary algorithms with neural networks have demonstrated foundational advancements in developing more effective and adaptive frameworks [22,55]. These studies highlight the feasibility of incorporating adaptive optimization techniques, which are important for managing the complexities and uncertainties inherent in peak load prediction.

Emerging trends also emphasize the integration of adaptive systems beyond traditional ML models. Hybrid models that combine ensemble learning with adaptive feature selection [20,238] are gaining increasing attention in research, offering improved robustness and flexibility in forecasting. Additionally, the utilization of anomaly detection methods within adaptive frameworks [21,149] allows models to identify and respond to irregular consumption patterns, further enhancing the forecasting accuracy. As research progresses, it is anticipated that the convergence of RL with these adaptive methods will promote more resilient and efficient peak demand forecasting systems, capable of leveraging real-time data and adapting to dynamic energy landscapes [64]. Moreover, graph neural networks are increasingly being explored for their ability to model topological and spatial relationships in power grids. By treating substations, feeders, and consumers as nodes in a graph, graph neural networks can capture spatial dependencies and interconnectivity patterns that influence the propagation of the peak load. While their application in peak demand forecasting remains nascent, recent studies on general load forecasting [256,257] suggest that graph neural networks hold promise for localized forecasting, especially in distributed grid environments.

6.2.4. Integration with smart grid technologies

IoT and real-time data, including smart meter readings and sensor networks, substantially improve the responsiveness of forecasts [204,207]. One optimal strategy is the deployment of edge-based preprocessing and local anomaly filtering, which reduces data transmission loads and enhances real-time responsiveness while preserving critical load characteristics. While interoperability across heterogeneous IoT systems remains a significant challenge, addressing it is crucial for enabling real-time forecasting that can directly support demand response strategies and grid operations. For example, smart meter data enable the disaggregation of household energy consumption, allowing models to identify specific patterns of appliance usage and predict peak demands with higher accuracy [85,228]. Additionally, the integration of sensor networks within industrial and commercial buildings facilitates the collection of detailed operational data, which can be leveraged to enhance STPDF [20,117]. To ensure scalability, cloud-based or modular streaming architectures allow real-time data pipelines to scale horizontally while maintaining low latency. Moreover, data quality can be improved through automated validation layers that filter out sensor faults, communication errors, and temporal inconsistencies prior to model ingestion. These advancements not only improve forecast precision but also support the integration of DERs such as solar PV and battery storage systems, thereby fostering a more resilient and sustainable energy grid [21,64].

6.2.5. Sustainability and resilience focus

Climate adaptation is increasingly shaping the objectives and methodologies of peak demand forecasting, particularly in the modeling of extreme events such as heatwaves or severe storms [29,48]. Models incorporating probabilistic methods or extreme value theory may yield better insights into risk and system resilience. By capturing the tail behavior of load distributions, these methods allow forecasters to quantify uncertainty and prepare for low-probability, high-impact events, which are expected to increase under climate change scenarios. As power systems incorporate larger proportions of renewables, these approaches also inform operational decisions about energy storage and distribution, ensuring grid stability even under climate-induced supply and demand fluctuations [21,207]. Recent studies show the growing emphasis on sustainability by combining probabilistic forecasting frameworks with policy-oriented variables, such as DSM participation and renewable energy penetration [42,66]. By simulating different policies or technology adoption scenarios, grid operators can evaluate possible resilience measures and identify high-risk periods well before they occur [125]. These methods can support adaptive planning by simulating how different external scenarios, such as policy shifts or environmental conditions, may affect system performance, thereby informing strategic decisions around flexible resource deployment.

Moreover, multi-resolution models capture both localized weather impacts and broader system trends, offering comprehensive insights into grid vulnerabilities and resource adequacy [199,238]. Furthermore, multi-horizon frameworks are gaining traction as they address overlapping forecasting needs across short-term, medium-term, and long-term planning horizons [8,225]. These comprehensive models integrate daily, weekly, and even annual projections to provide a unified platform for scheduling, operational control, and strategic investment decisions [157,207]. To further enhance climate resilience, forecasting frameworks should evolve toward integrated platforms that link probabilistic forecasting with scenario-based planning tools. This enables system operators to assess trade-offs between operational reliability, storage deployment, and emission reduction goals. These sustainability-focused efforts not only enhance the accuracy of predictions during normal conditions but also strengthen the ability of the grid to withstand and recover from extreme weather events, aligning peak demand forecasting with broader green energy goals and climate adaptation strategies [52,131].

7. Conclusion

This review of 186 studies traces the progression of peak demand forecasting from statistical baselines to machine learning and hybrid approaches, with emphasis on the elements that improve predictive performance in practice. It aggregates published evidence and does not include new case studies, proprietary methods, or interviews. These will be valuable in future work that traces organisational processes and deployment details in specific utilities. To retain contextual nuance without new data collection, we stratify the synthesis by forecast horizon, geographic scale, and data regime, and report areas of agreement and disagreement across study clusters. Our conclusions are therefore bounded by the literature base and are framed as patterns that recur across methods, horizons, and regions. Statistical models continue to play a central role because they make trends and seasonality explicit and provide transparent benchmarks. Machine learning methods add the capacity to capture non-linear effects and interactions that drive peaks when data are carefully curated and validation is rigorous. Forecasting outputs also need to move beyond a single peak value, since duration and variability determine reserves, ramping, and flexibility in systems with high renewable and distributed resources.

The choice of method is shaped by the operational task. Short-term peak forecasting benefits from sequential hybrids, where denoising or decomposition prepares the signal and non-linear learners capture residual spikes at sub-hourly or hourly horizons. Feeder and regional stability tasks are supported by parallel hybrids that combine complementary error patterns across locations while preserving local variability. Long-term planning requires physics-informed designs that incorporate capacity margins, demand response constraints, and tariff rules so that multi-year peak risks remain feasible under policy and network limits.

Machine learning has expanded the forecasting toolkit, but its use for peak events faces clear limitations that remain open for future research. Extremes are rare, which makes generalisation difficult and increases the likelihood of overfitting in complex hybrids. Smart meter and feeder data often contain gaps, noise, and misalignment, which can obscure peak shape and timing unless properly cleaned. Real-time applications are constrained by latency and computational load when forecasts must update continuously with new weather and distributed energy signals. Transparency also remains uneven. Post-hoc explanations provide partial insights, but operators and regulators require peak-focused justifications that identify the drivers of exceedance risk at a given lead time and clarify how uncertainty affects decisions such as reserve procurement or demand response.

From these observations, a focused roadmap for future research and practice can be outlined. Four priorities stand out. First, methods must become more robust by addressing data gaps, noise, and rare extreme events, supported by probabilistic forecasts that quantify uncertainty. Second, interpretability needs to be strengthened through attribution of peak risk to specific drivers and the use of intrinsically interpretable components. Third, scalability should be improved, for example, through edge preprocessing, quality gates, and federated learning that make high-frequency data usable without compromising privacy or timeliness. Finally, evaluation practices must be standardized, with stress-testing under missing data and backtesting against historic extremes reported consistently across horizons and regions.

Peak demand forecasting now has a broader toolset than a decade ago, but its value depends on how well methods are aligned with operational decisions. When hybrid strategies are matched to horizon and scale, when outputs capture both duration and variability, when uncertainty in peak risk is quantified, and when explanations are tailored to operator needs, machine learning approaches can improve accuracy and reliability while remaining usable for system operators, planners, and regulators.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors confirm that the data supporting this study are available within the article.

References

[1] S. Zhou, M. A. Brown, Smart meter deployment in europe: A comparative case study on the impacts of national policy schemes, Journal of cleaner production 144 (2017) 22–32.
[2] Ofgem, Smart meters in great britain, quarterly update june 2024, https://www.gov.uk/government/collections/smart-meters-statistics, 2024. Accessed on Nov 26 2024.
[3] H. Gharibvand, G. B. Gharehpetian, A. Anvari-Moghaddam, A survey on microgrid flexibility resources, evaluation metrics and energy storage effects, Renewable and Sustainable Energy Reviews 201 (2024) 114632.
[4] A. Sheffrin, H. Yoshimura, D. LaPlante, B. Neenan, Harnessing the power of demand, The Electricity Journal 21 (2008) 39–50.
[5] J. Saebi, H. Taheri, J. Mohammadi, S. S. Nayer, Demand bidding/buyback modeling and its impact on market clearing price, in: 2010 IEEE International Energy Conference, IEEE, 2010, pp. 791–796.
[6] P. Cappers, C. Goldman, D. Kathan, Demand response in us electricity markets: Empirical evidence, Energy 35 (2010) 1526–1535.
[7] C. Wang, P. Ju, F. Wu, X. Pan, Z. Wang, A systematic review on power system resilience from the perspective of generation, network, and load, Renewable and Sustainable Energy Reviews 167 (2022) 112567.
[8] J. Lee, Y. Cho, National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model?, Energy 239 (2022) 122366.
[9] J. Berrisch, M. Narajewski, F. Ziel, High-resolution peak demand estimation using generalized additive models and deep neural networks, Energy and AI 13 (2023) 100236.
[10] S. L. Koh, Y. S. Lim, Evaluating the economic benefits of peak load shifting for building owners and grid operator, in: 2015 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), IEEE, 2015, pp. 30–34.
[11] J. Sardi, N. Mithulananthan, M. Gallagher, D. Q. Hung, Multiple community energy storage planning in distribution networks using a cost-benefit analysis, Applied Energy 190 (2017) 453–463.
[12] H. Mao, X.-J. Zeng, G. Leng, Y.-J. Zhai, J. A. Keane, Short-term and midterm load forecasting using a bilevel optimization model, IEEE Transactions on Power Systems 24 (2009) 1080–1090.
[13] M. A. Alduailij, I. Petri, O. Rana, M. A. Alduailij, A. S. Aldawood, Forecasting peak energy demand for smart buildings, The Journal of Supercomputing 77 (2021) 6356–6380.
[14] M. Q. Raza, A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, Renewable and Sustainable Energy Reviews 50 (2015) 1352–1372.
[15] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, A. F. Adekoya, Electricity load forecasting: a systematic review, Journal of Electrical Systems and Information Technology 7 (2020) 1–19.
[16] S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, S. Aslam, A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids, Renewable and Sustainable Energy Reviews 144 (2021) 110992.
[17] Y. Eren, İ. Küçükdemiral, A comprehensive review on deep learning approaches for short-term load forecasting, Renewable and Sustainable Energy Reviews 189 (2024) 114031.
[18] S. Mohseni, A. C. Brent, S. Kelly, W. N. Browne, Demand response-integrated investment and operational planning of renewable and sustainable energy systems considering forecast uncertainties: A systematic review, Renewable and Sustainable Energy Reviews 158 (2022) 112095.
[19] J. Adinkrah, F. Kemausuor, E. T. Tchao, H. Nunoo-Mensah, A. S. Agbemenu, A. Adu-Poku, J. J. Kponyo, Artificial intelligence-based strategies for sustainable energy planning and electricity demand estimation: A systematic review, Renewable and Sustainable Energy Reviews 210 (2025) 115161.
[20] D.-H. Kim, E.-K. Lee, N. Qureshi, Peak-load forecasting for small industries: A machine learning approach, Sustainability 12 (2020) 6539.
[21] T. Zhang, X. Zhang, T. K. Chau, Y. Chow, T. Fernando, H. H.-C. Iu, et al., Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high pv penetration, Applied Energy 333 (2023) 120641.
[22] S. Ai, A. Chakravorty, C. Rong, Evolutionary ensemble lstm based household peak demand prediction, in: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE, 2019, pp. 1–6.
[23] M. A. Alduailij, I. Petri, O. Rana, M. A. Alduailij, A. S. Aldawood, Forecasting peak energy demand for smart buildings, The Journal of Supercomputing 77 (2021) 6356–6380.
[24] C.-L. Hor, S. J. Watson, S. Majithia, Daily load forecasting and maximum demand estimation using arima and garch, in: 2006 International conference on probabilistic methods applied to power systems, IEEE, 2006, pp. 1–6.
[25] N. Dongxiao, Z. Yunyun, L. Jinpeng, The application of time series seasonal multiplicative model and garch error amending model on forecasting the monthly peak load, in: 2009 International Forum on Computer Science-Technology and Applications, volume 3, IEEE, 2009, pp. 135–138.
[26] L. Saini, M. Soni, Artificial neural network based peak load forecasting using levenberg–marquardt and quasi-newton methods, IEE Proceedings-Generation, Transmission and Distribution 149 (2002) 578–584.
[27] M. Amin-Naseri, A. Soroush, Combined use of unsupervised and supervised learning for daily peak load forecasting, Energy Conversion and Management 49 (2008) 1302–1308.
[28] Z. Wang, Developed case-based reasoning system for short-term load forecasting, in: 2006 IEEE Power Engineering Society General Meeting, IEEE, 2006, pp. 1–6.
[29] Y. Li, B. Jones, The use of extreme value theory for forecasting long-term substation maximum electricity demand, IEEE Transactions on Power Systems PP (2019) 1.
[30] L. Jin, Y. J. Lai, T. X. Long, Peak load forecasting based on robust regression model, in: 2004 International Conference on Probabilistic Methods Applied to Power Systems, IEEE, 2004, pp. 123–128.
[31] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, PloS one 13 (2018) e0194889.
[32] V. Dobravec, N. Matak, C. Sakulin, G. Krajačić, Multilevel governance energy planning and policy: A view on local energy initiatives, Energy, Sustainability and Society 11 (2021) 1–17.
[33] H. Westskog, G. K. Hovelsrud, G. Sundqvist, How to make local context matter in national advice: Towards adaptive comanagement in norwegian climate adaptation, Weather, Climate, and Society 9 (2017) 267–283.
[34] C. Brandoni, F. Polonara, The role of municipal energy planning in the regional energy-planning process, Energy 48 (2012) 323–338.
[35] K. B. Debnath, M. Mourshed, Forecasting methods in energy planning models, Renewable and Sustainable Energy Reviews 88 (2018) 297–325.
[36] M. Hasan, Z. Mifta, S. J. Papiya, P. Roy, P. Dey, N. A. Salsabil, N.-U.-R. Chowdhury, O. Farrok, A state-of-the-art comparative review of load forecasting methods: Characteristics, perspectives, and applications, Energy Conversion and Management: X (2025) 100922.
[37] L. Suganthi, A. A. Samuel, Energy models for demand forecasting—a review, Renewable and sustainable energy reviews 16 (2012) 1223–1240.
[38] Z. Yu, Z. Niu, W. Tang, Q. Wu, Deep learning for daily peak load forecasting–a novel gated recurrent neural network combining dynamic time warping, IEEE Access 7 (2019) 17184–17194.
[39] N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Transactions on Power Systems 16 (2001) 798–805.
[40] E. Contaxi, C. Delkis, S. Kavatza, C. Vournas, The effect of humidity in a weather-sensitive peak load forecasting model, in: 2006 IEEE PES Power Systems Conference and Exposition, IEEE, 2006, pp. 1528–1534.
[41] R. Ramanathan, R. Engle, C. W. Granger, F. Vahid-Araghi, C. Brace, Short-run forecasts of electricity loads and peaks, International Journal of Forecasting 13 (1997) 161–174.
[42] Y. Jang, E. Byon, E. Jahani, K. Cetin, On the long-term density prediction of peak electricity load with demand side management in buildings, Energy and Buildings 228 (2020) 110450.
[43] S. Kim, H. Kim, A new metric of absolute percentage error for intermittent demand forecasts, International Journal of Forecasting 32 (2016) 669–679.
[44] K. Ho, Y.-Y. Hsu, C.-C. Yang, Short term load forecasting using a multilayer neural network with an adaptive learning algorithm, IEEE Transactions on Power Systems 7 (1992) 141–149.
[45] T. Onoda, Next day's peak load forecasting using an artificial neural network, in: [1993] Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems, IEEE, 1992, pp. 284–289.
[46] D. B. Belzer, M. A. Kellogg, Incorporating sources of uncertainty in forecasting peak power loads-a monte carlo analysis using the extreme value distribution, IEEE Transactions on Power Systems 8 (1993) 730–737.
[47] J.-G. Choi, J.-K. Park, K.-H. Kim, J.-C. Kim, A daily peak load forecasting system using a chaotic time series, in: Proceedings of International Conference on Intelligent System Application to Power Systems, IEEE, 1996, pp. 283–287.
[48] C. Sigauke, M. M. Nemukula, Modelling extreme peak electricity demand during a heatwave period: A case study, Energy systems 11 (2020) 139–161.

[49] T. Haida, S. Muto, Regression based peak load forecasting using a transformation technique, IEEE Transactions on Power Systems 9 (1994) 1788–1794.
[50] C. C. Hsu, C. Y. Chen, Regional load forecasting in taiwan—-applications of artificial neural networks, Energy Conversion & Management 44 (2003) 1941–1949.
[51] I. P. Panapakidis, G. C. Christoforidis, N. Asimopoulos, A. S. Dagoumas, Combining wavelet transform and support vector regression model for day-ahead peak load forecasting in the greek power system, in: 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), IEEE, 2017, pp. 1–6.
[52] Y. Chen, C. Obrecht, F. Kuznik, Enhancing peak prediction in residential load forecasting with soft dynamic time wrapping loss functions, Integrated Computer-Aided Engineering (2024) 1–14.
[53] V. Mtembo, G. A. Taylor, A. Ekwue, A novel econometric model for peak demand forecasting, in: 2014 49th International Universities Power Engineering Conference (UPEC), IEEE, 2014, pp. 1–6.
[54] E. Khorsheed, Long-term energy peak load forecasting models: A hybrid statistical approach, in: 2018 Advances in Science and Engineering Technology International Conferences (ASET), IEEE, 2018, pp. 1–6.
[55] J. Xue, Z. Xu, J. Watada, Building an integrated hybrid model for short-term and mid-term load forecasting with genetic optimization, International journal of innovative computing, information & control: IJICIC 8 (2012) 7381–7391.
[56] J. Liu, L. E. Brown, Prediction of hour of coincident daily peak load, in: 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE, 2019, pp. 1–5.
[57] H.-T. Yang, T.-C. Liang, K.-R. Shih, C.-L. Huang, Power system yearly peak load forecasting: a grey system modeling approach, in: Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD'95, volume 1, IEEE, 1995, pp. 261–266.
[58] K. Gajowniczek, R. Nafkha, T. Ząbkowski, Electricity peak demand classification with artificial neural networks, in: 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2017, pp. 307–315.
[59] K. Gajowniczek, R. Nafkha, T. Ząbkowski, Seasonal peak demand classification with machine learning techniques, in: 2018 International Conference on Applied Mathematics & Computer Science (ICAMCS), IEEE, 2018, pp. 101–1014.
[60] J. Liu, L. E. Brown, Effect of forecast accuracy on day ahead prediction of coincident peak days, in: 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), IEEE, 2019, pp. 661–666.
[61] P. E. Mcsharry, S. Bouwman, G. Bloemhof, Probabilistic forecasts of the magnitude and timing of peak electricity demand, IEEE Transactions on Power Systems (2005) 1166–1172.
[62] Y. K. Bichpuriya, S. Soman, A. Subramanyam, Robust probability density forecasts of yearly peak load using non-parametric model, in: 2016 IEEE Power and Energy Society General Meeting (PESGM), IEEE, 2016, pp. 1–5.
[63] Y. He, C. Cao, J. Xiao, Day-ahead peak load probability density forecasting based on qrlstm-df considering exogenous factors, IEEE Transactions on Industrial Informatics PP (2022) 1–10.
[64] Y. He, Y. Liu, W. Zhang, Probability density prediction of peak load based on mixed frequency noise-assisted multivariate empirical mode decomposition, Applied Intelligence 54 (2024) 2648–2672.
[65] J. E. Boylan, A. A. Syntetos, Intermittent demand forecasting: Context, methods and applications, John Wiley & Sons, 2021.
[66] N. L. Rajakovic, V. M. Shiljkut, Long-term forecasting of annual peak load considering effects of demand-side programs, Journal of Modern Power Systems and Clean Energy 6 (2018) 145–157.
[67] D. Gillies, B. Bernholtz, P. Sandiford, A new approach to forecasting daily peak loads [includes discussion], Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems 75 (1956) 382–387.
[68] L. E. Stetson, G. L. Stark, Peak electrical demands of individuals and groups of rural residential customers, IEEE Transactions on Industry Applications 24 (1988) 772–776.
[69] A. D. Papalexopoulos, A regression-based approach to short-term system load forecasting, IEEE Transactions on Power Systems 5 (1990) 1535–1547.
[70] R. P. Broadwater, A. Sargent, Estimating substation peaks from load research data, IEEE Transactions on Power Delivery 12 (1997) 451–456.
[71] L. Jin, Y. Jilai, L. Zhuo, Load forecast for system operation in peak load period, in: 2007 IEEE Power Engineering Society General Meeting, IEEE, 2007, pp. 1–6.
[72] A. Goia, C. May, G. Fusai, Functional clustering and linear regression for peak load forecasting, International Journal of Forecasting 26 (2010) 700–711.
[73] J. Huang, Y. Li, Y. Liu, Summer daily peak load forecasting considering accumulation effect and abrupt change of temperature, in: 2012 IEEE Power and Energy Society General Meeting, IEEE, 2012, pp. 1–4.
[74] J. Moon, S. Park, S. Rho, E. Hwang, Interpretable short-term electrical load forecasting scheme using cubist, Computational intelligence and neuroscience 2022 (2022) 6892995.
[75] G.-C. Lee, Regression-based methods for daily peak load forecasting in south korea, Sustainability 14 (2022) 3984.
[76] S. Towill, Estimation of maximum demand on a british electricity-board system. forecast periods of 1–3 years, in: Proceedings of the Institution of Electrical Engineers, volume 121, IET, 1974, pp. 609–615.
[77] H. K. Temraz, K. M. El-Nagar, M. M. A. Salama, Applications of noniterative least absolute value estimation for forecasting annual peak electric power demand, Canadian Journal of Electrical and Computer Engineering 23 (1998) 141–146.
[78] S. M. Maksimovich, V. M. Shiljkut, The peak load forecasting afterwards its intensive reduction, IEEE transactions on power delivery 24 (2009) 1552–1559.
[79] R. J. Hyndman, S. Fan, Density forecasting for long-term peak electricity demand, IEEE Transactions on Power Systems 25 (2009) 1142–1153.
[80] J. Steinfeld, A. Bruce, M. Watt, Peak load characteristics of sydney office buildings and policy recommendations for peak load reduction, Energy and Buildings 43 (2011) 2179–2187.
[81] Z. El-Razaz, N. Al-Mohawes, Weekly peak load forecasting for fast-developing cities, Canadian Electrical Engineering Journal 11 (1986) 184–187.
[82] F. A. Razak, A. H. Hashim, I. Z. Abidin, M. Shitan, Malaysian peak daily load forecasting, in: 2009 IEEE Student Conference on Research and Development (SCOReD), IEEE, 2009, pp. 392–394.
[83] Z. Yang, Modelling and forecasting of electric daily peak load movement based on the elliptic-orbit model with weekly periodic extension: a case study, IET Generation, Transmission & Distribution 8 (2014) 2046–2054.
[84] A. Ananthasingam, A. Atputharajah, Forecast daily night peak electric power demand in sri lankan power system, in: 2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS), IEEE, 2015, pp. 238–243.
[85] B. Nepal, M. Yamaha, A. Yokoe, T. Yamaji, Electricity load forecasting using clustering and arima model for energy management in buildings, Japan Architectural Review 3 (2020) 62–76.
[86] P. Gupta, A stochastic approach to peak power-demand forecasting in electric utility systems, IEEE Transactions on Power Apparatus and Systems (1971) 824–832.
[87] K. N. Stanton, Medium-range, weekly and seasonal peak demand forecasting by probability methods, IEEE Transactions on Power Apparatus and Systems (1971) 1183–1189.
[88] Y. Kareem, A. R. Majeed, Monthly peak-load demand forecasting for sulaimany governorate using sarima., in: 2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America, IEEE, 2006, pp. 1–5.
[89] S. R. Rallapalli, S. Ghosh, Forecasting monthly peak demand of electricity in india—a critique, Energy Policy 45 (2012) 516–520.
[90] K. N. Stanton, P. C. Gupta, Forecasting annual or seasonal peak demand in electric utility systems, IEEE Transactions on Power Apparatus & Systems PAS-89 (1970) 951–959.
[91] P. E. McSharry, S. Bouwman, G. Bloemhof, Probabilistic forecasts of the magnitude and timing of peak electricity demand, IEEE Transactions on Power Systems 20 (2005) 1166–1172.
[92] Y. K. Bichpuriya, S. Soman, A. Subramanyam, Non-parametric probability density forecast of an hourly peak load during a month, in: 2014 Power Systems Computation Conference, IEEE, 2014, pp. 1–6.
[93] S. Nuchprayoon, Forecasting of daily load curve on monthly peak day using load research data and harmonics model, in: 2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), IEEE, 2016, pp. 338–342.
[94] M. O. M. Mahmoud, F. Mhamdi, M. Jaidane-Saidane, Long term multi-scale analysis of the daily peak load based on the empirical mode decomposition, in: 2009 IEEE Bucharest PowerTech, IEEE, 2009, pp. 1–6.
[95] Y.-Y. Hsu, C.-C. Yang, Design of artificial neural networks for short-term load forecasting. part 2: Multilayer feedforward networks for peak load and valley load forecasting, in: IEE Proceedings C (Generation, Transmission and Distribution), volume 138, IET, 1991, pp. 414–418.
[96] Y.-Y. Hsu, C.-C. Yang, Design of artificial neural networks for short-term load forecasting. part 1: Self-organising feature maps for day type identification, in: IEE Proceedings C (Generation, Transmission and Distribution), volume 138, IET, 1991, pp. 407–413.
[97] D. C. Park, O. Mohammed, Artificial neural network based electric peak load forecasting, in: IEEE Proceedings of the SOUTHEASTCON'91, IEEE, 1991, pp. 225–228.
[98] S. Saeed Madani, Electric load forecasting using an artificial neural network, IEEE Transactions on Power Systems 6 (1991) 442–449.
[99] E. Barakat, J. Al-Qassim, S. Al Rashed, New model for peak demand forecasting applied to highly complex load characteristics of a fast developing area, in: IEE Proceedings C (Generation, Transmission and Distribution), volume 139, IET, 1992, pp. 136–140.
[100] Y. Morioka, K. Sakurai, A. Yokoyama, Y. Sekine, Next day peak load forecasting using a multilayer neural network with an additional learning, in: [1993] Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems, IEEE, 1992, pp. 60–65.
[101] Y. Mizukami, T. Nishimori, Maximum electric power demand prediction by neural network, in: [1993] Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems, IEEE, 1992, pp. 296–301.
[102] A. AlFuhaid, M. El-Sayed, M. Mahmoud, Cascaded artificial neural networks for short-term load forecasting, IEEE Transactions on Power Systems 12 (1997) 1524–1529.
[103] M. A. Aboul-Magd, et al., An artificial neural network model for electrical daily peak load forecasting with an adjustment for holidays, in: LESCOPE 01. 2001 Large Engineering Systems Conference on Power Engineering. Conference Proceedings. Theme: Powering Beyond 2001 (Cat. No. 01ex490), IEEE, 2001, pp. 105–113.
[104] T. Matsui, T. Iizaka, Y. Fukuyama, Peak load forecasting using analyzable structured neural network, in: 2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 01CH37194), volume 2, IEEE, 2001, pp. 405–410.
[105] L. M. Saini, M. K. Soni, Artificial neural network-based peak load forecasting using conjugate gradient methods, IEEE Transactions on Power Systems 17 (2002) 907–912.
[106] T. Yalcinoz, U. Eminoglu, Short term and medium term power distribution load forecasting by neural networks, Energy Conversion and Management 46 (2005) 1393–1405.

[107] Z. Wang, Y. Cao, Mutual information and non-fixed anns for daily peak load forecasting, in: 2006 IEEE PES Power Systems Conference and Exposition, IEEE, 2006, pp. 1523–1527.
[108] L. Jin, Y. Feng, Y. Jilai, Peak load forecasting using hierarchical clustering and rprop neural network, in: 2006 IEEE PES Power Systems Conference and Exposition, IEEE, 2006, pp. 1535–1540.
[109] M. Ghomi, M. Goodarzi, M. Goodarzi, Peak load forecasting of electric utilities for west province of iran by using neural network without weather information, in: 2010 12th International Conference on Computer Modelling and Simulation, IEEE, 2010, pp. 28–32.
[110] L. M. Saini, Peak load forecasting using bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks, Electric Power Systems Research 78 (2008) 1302–1310.
[111] M. B. Tasre, P. P. Bedekar, V. N. Ghate, Daily peak load forecasting using ann, in: 2011 Nirma University International Conference on Engineering, IEEE, 2011, pp. 1–6.
[112] J. Milojković, I. Litovski, V. Litovski, Ann application for the next day peak electricity load prediction, in: 11th Symposium on Neural Network Applications in Electrical Engineering, IEEE, 2012, pp. 237–241.
[113] D. Abdellah, L. Djamel, Forecasting the algerian load peak profile using time series model based on backpropagation neural networks, in: 4th International Conference on Power Engineering, Energy and Electrical Drives, IEEE, 2013, pp. 1734–1737.
[114] L. Hernández, C. Baladón, J. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, J. Sanjuán, á. González, J. Lloret, Improved short-term load forecasting based on two-stage predictions with artificial neural networks in a microgrid environment, Energies 6 (2013) 4489–4507.
[115] S. Chemetova, P. Santos, M. Ventim-Neves, Load peak forecasting in different load patterns situations, in: 2016 10th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), IEEE, 2016, pp. 148–151.
[116] Y. T. Chae, R. Horesh, Y. Hwang, Y. M. Lee, Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings, Energy & Buildings 111 (2016) 184–194.
[117] M. H. Leak, G. K. Venayagamoorthy, Forecasting peak daily load in distribution feeders, in: 2018 Clemson University Power Systems Conference (PSC), IEEE, 2018, pp. 1–8.
[118] S. Chemetova, P. Santos, A. J. Pires, Peak load forecasting in electrical deregulated market environment-the dynamic tariffs, in: IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, volume 1, IEEE, 2019, pp. 2227–2232.
[119] A. Ganguly, A. K. Sil, Development of regional load management system based on rural, semi urban and urban loads-a critical analysis, Bulletin of Electrical Engineering and Informatics 11 (2022) 68–81.
[120] N. Kebir, A. Lamallam, A. Moussa, Daily peak-based short-term demand prediction using backpropagation combined to chi-squared distribution, International Journal of Emerging Electric Power Systems 21 (2020) 20200098.
[121] J. Jang, W. Jeong, S. Kim, B. Lee, M. Lee, J. Moon, Raid: Robust and interpretable daily peak load forecasting via multiple deep neural networks and shapley values, Sustainability 15 (2023) 6951.
[122] M. Saglam, X. Lv, C. Spataru, O. A. Karaman, Instantaneous electricity peak load forecasting using optimization and machine learning, Energies 17 (2024) 777.
[123] W. Ali, M. Hussein, L. Mina, Monthly peak-load demand forecasting for sulaimani governorate using different weather conditions based on ann model, SULAIMANI JOURNAL FOR ENGINEERING SCIENCES 7 (2020) 11–22.
[124] A. Sharma, S. K. Jain, A novel two-stage framework for mid-term electric load forecasting, IEEE Transactions on Industrial Informatics 20 (2023) 247–255.
[125] B. Heidrich, M. Hertel, O. Neumann, V. Hagenmeyer, R. Mikut, Using conditional invertible neural networks to perform mid-term peak load forecasting, IET Smart Grid (2024).
[126] K. Nagasaka, M. Al Mamun, Long-term peak demand prediction of 9 japanese power utilities using radial basis function networks, in: IEEE Power Engineering Society General Meeting, 2004., IEEE, 2004, pp. 315–322.
[127] S. Phimphachan, K. Chamnongthai, P. Kumhom, N. Jittiwarangkul, A. Sangswang, Energy and peak load forecast models using neural network for fast developing area, in: IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004., volume 1, IEEE, 2004, pp. 389–393.
[128] S. Phimphachanh, K. Chammongthai, P. Kumhom, A. Sangswang, Using neural network for long term peak load forecasting in vientiane municipality, in: 2004 IEEE Region 10 Conference TENCON 2004., volume 100, IEEE, 2004, pp. 319–322.
[129] O. A. S. Carpinteiro, R. C. Leme, A. C. Z. D. Souza, C. A. M. Pinheiro, E. M. Moreira, Long-term load forecasting via a hierarchical neural model with time integrators, Electric Power Systems Research 77 (2006) 371–378.
[130] L. Ghods, M. Kalantar, Long-term peak demand forecasting by using radial basis function neural networks, Iranian Journal of Electrical and Electronic Engineering 6 (2010) 175–182.
[131] C. Effiong, O. Simeon, F. O. Faithpraise, Modelling and forecasting peak load demand in uyo metropolis using artificial neural network technique, Development 7 (2020).
[132] J. Faraji, H. Hashemi-Dezaki, A. Ketabi, Multi-year load growth-based optimal planning of grid-connected microgrid considering long-term load demand forecasting: A case study of tehran, iran, Sustainable Energy Technologies and Assessments 42 (2020) 100827.
[133] J. Yu, H. Lee, Y. Jeong, S. Kim, Linguistic fuzzy modeling approach for daily peak load forecasting, in: 2013 International Conference on Fuzzy Theory and Its Applications (iFUZZY), IEEE, 2013, pp. 116–121.
[134] J. Jamaaluddin, D. Hadidjaja, I. Sulistiyowati, E. Suprayitno, I. Anshory, S. Syahrorini, Very short term load forecasting peak load time using fuzzy logic, in: IOP Conference Series: Materials Science and Engineering, volume 403, IOP Publishing, 2018, p. 012070.
[135] Y. T. Nugraha, P. M. Simanjuntak, M. Irwanto, R. Rida, M. Othman, et al., Analysis of forecast of renewable energy development in north sumatra using anfis, Jurnal Media Elektro (2024) 27–36.
[136] S. Kato, K. Yukita, Y. Goto, K. Ichiyanagi, Study of daily peak load forecasting by structured representation on genetic algorithms for function fitting, in: IEEE/PES Transmission and Distribution Conference and Exhibition, volume 3, IEEE, 2002, pp. 1686–1690.
[137] P. T. T. Nguyen, L. Manuel, City-scale electricity demand forecasting using a gaussian process model, in: 2020 5th International Conference on Green Technology and Sustainable Development (GTSD), 2020, pp. 405–412. doi:https://doi.org/10.1109/GTSD50082.2020.9303132.
[138] M. N. S. K. Shabbir, M. Z. Ali, X. Liang, M. S. A. Chowdhury, A probabilistic approach considering contingency parameters for peak load demand forecasting, Canadian Journal of Electrical and Computer Engineering 41 (2018) 224–233.
[139] P. Atsawathawichok, P. Teekaput, T. Ploysuwan, Long term peak load forecasting in thailand using multiple kernel gaussian process, in: 2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE, 2014, pp. 1–4.
[140] T. Ploysuwan, P. Atsawathawichok, P. Teekaput, Peak load forecasting of electricity generating authority of thailand by gaussian process, in: 2014 International Electrical Engineering Congress (iEECON), IEEE, 2014, pp. 1–4.
[141] J.-Z. Wang, L. Wu, H.-Y. Lu, Special periods peak load analysis and superior forecasting method based on ls-svm, in: 2007 International Conference on Wavelet Analysis and Pattern Recognition, volume 1, IEEE, 2007, pp. 249–253.
[142] E. El-Attar, J. Goulermas, Q. Wu, Forecasting electric daily peak load based on local prediction, in: 2009 IEEE Power & Energy Society General Meeting, IEEE, 2009, pp. 1–6.
[143] J. Dhillon, S. A. Rahman, S. U. Ahmad, M. J. Hossain, Peak electricity load forecasting using online support vector regression, in: 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE, 2016, pp. 1–4.
[144] P. Zeng, M. Jin, Peak load forecasting based on multi-source data and day-to-day topological network, IET Generation, Transmission Distribution 12 (2018) 1374–1381.
[145] M. K. Azad, S. Uddin, M. Takruri, Support vector regression based electricity peak load forecasting, in: 2018 11th international symposium on mechatronics and its applications (ISMA), IEEE, 2018, pp. 1–5.
[146] H. Komatsu, O. Kimura, Peak demand alert system based on electricity demand forecasting for smart meter data, Energy and Buildings 225 (2020) 110307.
[147] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Applied Energy 127 (2014) 1–10.
[148] X. Deng, A. Ye, J. Zhong, D. Xu, W. Yang, Z. Song, Z. Zhang, J. Guo, T. Wang, Y. Tian, H. Pan, Z. Zhang, H. Wang, C. Wu, J. Shao, X. Chen, Bagging–xgboost algorithm based extreme weather identification and short-term load forecasting model, Energy Reports 8 (2022) 8661–8674.
[149] O. Laayati, M. Bouzi, A. Chebak, Smart energy management system: Design of a monitoring and peak load forecasting system for an experimental open-pit mine, Applied System Innovation 5 (2022) 1–18.
[150] T. Fu, H. Zhou, X. Ma, Z. J. Hou, D. Wu, Predicting peak day and peak hour of electricity demand with ensemble machine learning, 2022. https://arxiv.org/abs/2203.13886. arXiv:2203.13886.
[151] L. Semmelmann, O. Resch, S. Henni, C. Weinhardt, Privacy-preserving peak time forecasting with learning to rank xgboost and extensive feature engineering, IET Smart Grid 7 (2023) n/a–n/a.
[152] I. Pahendra, E. Darnila, M. Sadli, M. Sinambela, W. Fuadi, et al., Peak load forecasting based on long short term memory, in: 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), IEEE, 2019, pp. 137–140.
[153] I. A. Ibrahim, M. J. Hossain, Lstm neural network model for ultra-short-term distribution zone substation peak demand prediction, in: 2020 IEEE Power & Energy Society General Meeting (PESGM), 2020, pp. 1–5. doi:https://doi.org/10.1109/PESGM41954.2020.9281973.
[154] A. Soman, A. Trivedi, D. Irwin, B. Kosanovic, B. McDaniel, P. Shenoy, Peak forecasting for battery-based energy optimizations in campus microgrids, in: Proceedings of the Eleventh ACM International Conference on Future Energy Systems, e-Energy '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 237–241. URL: https://doi.org/10.1145/3396851.3397751.
[155] N. Mughees, S. A. Mohsin, A. Mughees, A. Mughees, Deep sequence to sequence bi-lstm neural networks for day-ahead peak load forecasting, Expert Systems with Applications 175 (2021) 114844.
[156] K. Zhu, Y. Li, W. Mao, F. Li, J. Yan, Lstm enhanced by dual-attention-based encoder-decoder for daily peak load forecasting, Electric Power Systems Research 208 (2022) 107860.
[157] B.-S. Kwon, R.-J. Park, K.-B. Song, Weekly peak load forecasting for 104 weeks using deep learning algorithm, in: 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), IEEE, 2019, pp. 1–4.
[158] E. Barakat, M. Eissa, Forecasting monthly peak demand in fast growing electric utility using a composite multiregression-decomposition model, in: IEE Proceedings C (Generation, Transmission and Distribution), volume 136, IET, 1989, pp. 35–41.
[159] K.-L. Ho, Y.-Y. Hsu, C.-F. Chen, T.-E. Lee, C.-C. Liang, T.-S. Lai, K.-K. Chen, Short term load forecasting of taiwan power system using a knowledge-based expert system, IEEE Transactions on Power Systems 5 (1990) 1214–1221.
[160] I. Drezga, S. Rahman, Short-term load forecasting with local ann predictors, IEEE Transactions on Power Systems 14 (1999) 844–850.

[161] A. Fadhilah, S. Suriawati, H. Amir, Z. Izham, S. Mahendran, Malaysian day-type load forecasting, in: 2009 3rd International Conference on Energy and Environment (ICEE), IEEE, 2009, pp. 408–411.
[162] S. Kiartzis, A. Bakirtzis, J. Theocharis, G. Tsagas, A fuzzy expert system for peak load forecasting. application to the greek power system, in: 2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No. 00CH37099), volume 3, IEEE, 2000, pp. 1097–1100.
[163] M.-R. Lee, S.-J. Wang, L. Yi-Yu, L.-I. Tai, H.-J. Shi, The maximum power demand forecasting with fuzzy theory, in: 2010 International Symposium on Computer, Communication, Control and Automation (3CA), volume 2, IEEE, 2010, pp. 419–422.
[164] X. Zhang, T. K. Chau, Y. H. Chow, T. Fernando, H. H.-C. Iu, A novel sequence to sequence data modelling based cnn-lstm algorithm for three years ahead monthly peak load forecasting, IEEE Transactions on Power Systems 39 (2023) 1932–1947.
[165] Z. S. Elrazaz, A. A. Mazi, Unified weekly peak load forecasting for fast growing power system, IEE proceedings. Part C, Generation, transmission and distribution 136 (1989) 29.
[166] E. Barakat, M. Qayyum, M. Hamed, S. Al Rashed, Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics. i. application of classical time-series methods. ii. improved modelling of system dynamic load characteristics, IEEE Transactions on Power Systems 5 (1990) 813–824.
[167] F. A. Razak, A. H. Hashim, I. Z. Abidin, M. Shitan, Moving holidays' effects on the malaysian peak daily load, in: 2010 IEEE International Conference on Power and Energy, IEEE, 2010, pp. 906–910.
[168] Y.-Y. Hsu, K.-L. Ho, Fuzzy expert systems: an application to short-term load forecasting, in: IEE Proceedings C (Generation, Transmission and Distribution), volume 139, IET, 1992, pp. 471–477.
[169] T. Haida, S. Muto, Regression based peak load forecasting using a transformation technique, IEEE Transactions on Power Systems 9 (1994) 1788–1794.
[170] T. Haida, S. Muto, Y. Takahashi, Y. Ishi, Peak load forecasting using multiple-year data with trend data processing techniques, Electrical Engineering in Japan 124 (1998) 7–16.
[171] H. Alfares, M. Nazeeruddin, Regression-based methodology for daily peak load forecasting, in: Proceedings of the 2nd International Conference on Operations and Quantitative Management, volume 3, 1999, pp. 468–471.
[172] P. Dash, H. Satpathy, S. Rahman, Short term daily average and peak load predications using a hybrid intelligent approach, in: Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD'95, volume 2, IEEE, 1995, pp. 565–570.
[173] S. Mandal, A. Agrawal, Fuzzy-neural network based short term peak and average load forecasting (stpalf) system with network security, in: IECEC-97 Proceedings of the Thirty-Second Intersociety Energy Conversion Engineering Conference (Cat. No. 97CH6203), volume 3, IEEE, 1997, pp. 2193–2196.
[174] M. Rashidi, F. Rashidi, H. Monavar, Peak load forecasting in power systems using emotional learning based fuzzy logic, in: SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483), volume 2, IEEE, 2003, pp. 1985–1988.
[175] L. Jin, L. Ziyang, S. Jingbo, S. Xinying, An efficient method for peak load forecasting, in: 2005 International Power Engineering Conference, IEEE, 2005, pp. 1–6.
[176] M. A. Badri, A. Al-Mutawa, D. Davis, D. Davis, Edssf: A decision support system (dss) for electricity peak-load forecasting, Energy 22 (1997) 579–589.
[177] J. Nazarko, W. Zalewski, The fuzzy regression approach to peak load estimation in power distribution systems, IEEE Transactions on Power Systems 14 (1999) 809–814.
[178] G. Cartina, V. Alexandrescu, G. Grigoras, M. Moshe, Peak load estimation in distribution networks by fuzzy regression approach, in: 2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No. 00CH37099), volume 3, IEEE, 2000, pp. 907–910.
[179] M. S. Kandil, S. M. El-Debeiky, N. E. Hasanien, Long-term load forecasting for fast developing utility using a knowledge-based expert system, IEEE Power Engineering Review 22 (2002) 78.
[180] O. A. Carpinteiro, R. C. Leme, A. de Souza, et al., A hierarchical hybrid neural model with time integrators in long-term peak-load forecasting, in: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., volume 5, IEEE, 2005, pp. 2960–2965.
[181] M. Gavrilas, C. V. Sfintes, O. Ivanov, Comparison of neural and evolutionary approaches to peak load estimation in distribution systems, in: EUROCON 2005-The International Conference on" Computer as a Tool", volume 2, IEEE, 2005, pp. 1461–1464.
[182] L. Ran, G. Chaoyun, The relevance analysis between electrical day peak load and meteorological index based on wavelet denoising and svm, in: 2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, IEEE, 2008, pp. 788–793.
[183] J. Zhengyuan, Electricity consumption forecasting in peak load month based on variable weight combination forecasting model, in: 2008 IEEE International Conference on Automation and Logistics, IEEE, 2008, pp. 1265–1269.
[184] M. Beiraghi, A. Ranjbar, Discrete fourier transform based approach to forecast monthly peak load, in: 2011 Asia-Pacific Power and Energy Engineering Conference, IEEE, 2011, pp. 1–5.
[185] G. M. Khan, S. Khan, F. Ullah, Short-term daily peak load forecasting using fast learning neural network, in: 2011 11th International Conference on Intelligent Systems Design and Applications, IEEE, 2011, pp. 843–848.
[186] M. Moazzami, A. Khodabakhshian, R. Hooshmand, A new hybrid day-ahead peak load forecasting method for iran's national grid, Applied Energy 101 (2013) 489–501.
[187] O. Ivanov, M. Gavrilaş, Multilayer perceptron architecture optimization for peak load estimation, in: 12th Symposium on neural network applications in electrical engineering (NEUREL), IEEE, 2014, pp. 67–72.
[188] S. Takiyar, Grid reliability enhancement by peak load forecasting with a pso hybridized ann model, in: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), IEEE, 2015, pp. 1–6.
[189] A. Jarndal, S. Hamdan, Forecasting of peak electricity demand using annga and ann-pso approaches, in: 2017 7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO), IEEE, 2017, pp. 1–5.
[190] Y. Huang, N. Hasan, C. Deng, Y. Bao, Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting, Energy 239 (2022) 122245.
[191] K. M. El-naggar, K. A. Al-rumaih, Electric load forecasting using genetic based algorithm, optimal filter estimator and least error squares technique: Comparative study, World Academy of Science, Engineering and Technology 6 (2005) 138–142.
[192] R. Torkzadeh, A. Mirzaei, M. M. Mirjalili, A. S. Anaraki, M. R. Sehhati, F. Behdad, Medium term load forecasting in distribution systems based on multi linear regression & principal component analysis: A novel approach, in: 2014 19th Conference on Electrical Power Distribution Networks (EPDC), IEEE, 2014, pp. 66–70.
[193] L. Aswanuwath, W. Pannakkong, J. Buddhakulsomsiri, J. Karnjana, V.-N. Huynh, A hybrid model of vmd-emd-fft, similar days selection method, stepwise regression, and artificial neural network for daily electricity peak load forecasting, Energies 16 (2023) 1860.
[194] Z. Tavassoli-Hojati, S. Ghaderi, H. Iranmanesh, P. Hilber, E. Shayesteh, A self-partitioning local neuro fuzzy model for short-term load forecasting in smart grids, Energy 199 (2020) 117514.
[195] C. Heylman, Y. G. Kim, J. Wang, Forecasting energy trends and peak usage at the university of virginia, in: 2015 Systems and Information Engineering Design Symposium, IEEE, 2015, pp. 362–368.
[196] A. Laouafi, M. Mordjaoui, F. Laouafi, T. E. Boukelia, Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology, International Journal of Electrical Power & Energy Systems 77 (2016) 136–144.
[197] J. R. G. Sarduy, K. G. D. Santo, M. A. Saidel, Linear and non-linear methods for prediction of peak load at university of são paulo, Measurement 78 (2016) 187–201.
[198] K. Ogihara, S. Urano, A study of risk reduction for daily peak load demand forecasting, in: 2019 IEEE Milan PowerTech, IEEE, 2019, pp. 1–6.
[199] J. Berrisch, M. Narajewski, F. Ziel, High-resolution peak demand estimation using generalized additive models and deep neural networks, Energy and AI 13 (2023) 100236.
[200] M. Zulfiqar, M. Kamran, M. Rasheed, T. Alquthami, A. Milyani, Hyperparameter optimization of support vector machine using adaptive differential evolution for electricity load forecasting, Energy Reports 8 (2022) 13333–13352.
[201] C. Tairen, J. M. Lehr, L. Olga, M.-R. Manel, Distribution feeder-level day-ahead peak load forecasting methods and comparative study, IET Generation Transmission & Distribution 12 (2018) 3270–3278.
[202] C. Guo, Q. Ge, H. Jiang, G. Yao, Q. Hua, Maximum power demand prediction using fbprophet with adaptive kalman filtering, IEEE Access 8 (2020) 19236–19247.
[203] M.-R. Kazemzadeh, A. Amjadian, T. Amraee, A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting, Energy 204 (2020).
[204] E. U. Haq, X. Lyu, Y. Jia, M. Hua, F. Ahmad, Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach, Energy Reports 6 (2020) 1099–1105.
[205] A. Almazrouee, A. Almeshal, A. Almutairi, M. Alenezi, S. Alhajeri, Long-term forecasting of electrical loads in kuwait using prophet and holt-winters models, Applied Sciences 10 (2020) 5627.
[206] B. Jin, G. Zeng, Z. Lu, H. Peng, S. Luo, X. Yang, H. Zhu, M. Liu, Hybrid lstm–bpnn-to-bpnn model considering multi-source information for forecasting medium- and long-term electricity peak load, Energies 15 (2022) 7584.
[207] W. Waheed, Q. Xu, Data-driven short term load forecasting with deep neural networks: Unlocking insights for sustainable energy management, Electric Power Systems Research 232 (2024) 110376.
[208] S. Hosseini, B. M. H. Zade, New hybrid method for attack detection using combination of evolutionary algorithms, svm, and ann, Computer Networks 173 (2020) 107168.
[209] K. S. Garud, S. Jayaraj, M.-Y. Lee, A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models, International Journal of Energy Research 45 (2021) 6–35.
[210] J. Efstathiou, Expert systems, fuzzy logic and rule-based control explained at last, Transactions of the Institute of Measurement and Control 10 (1988) 198–206.
[211] Z. Cao, X. Han, W. Lyons, F. O'Rourke, Energy management optimisation using a combined long short-term memory recurrent neural network–particle swarm optimisation model, Journal of Cleaner Production 326 (2021) 129246.
[212] B. Predić, L. Jovanovic, V. Simic, N. Bacanin, M. Zivkovic, P. Spalevic, N. Budimirovic, M. Dobrojevic, Cloud-load forecasting via decomposition-aided attention recurrent neural network tuned by modified particle swarm optimization, Complex & Intelligent Systems 10 (2024) 2249–2269.

[213] Y. Huang, C. Li, Accurate heating, ventilation and air conditioning system load prediction for residential buildings using improved ant colony optimization and wavelet neural network, Journal of Building Engineering 35 (2021) 101972.
[214] F. R. Hampel, Introduction to Huber (1964) Robust Estimation of a Location Parameter, Springer New York, New York, NY, 1992, pp. 479–491.
[215] S. Feuerriegel, P. Bodenbenner, D. Neumann, Value and granularity of ict and smart meter data in demand response systems, Energy Economics 54 (2016) 1–10.
[216] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural networks 2 (1989) 359–366.
[217] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research 15 (2014) 1929–1958.
[218] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (1995) 273–297.
[219] L. Breiman, Bagging predictors, Machine learning 24 (1996) 123–140.
[220] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.
[221] R. Polikar, Ensemble learning, in: Ensemble machine learning: Methods and applications, Springer, 2012, pp. 1–34.
[222] T. Ahmad, H. Chen, Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment, Energy 160 (2018) 1008–1020.
[223] N. Zhang, Z. Li, X. Zou, S. M. Quiring, Comparison of three short-term load forecast models in southern california, Energy 189 (2019) 116358.
[224] H. Lu, F. Cheng, X. Ma, G. Hu, Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: a case study of an intake tower, Energy 203 (2020) 117756.
[225] P. Bovornkeeratiroj, J. Wamburu, D. Irwin, P. Shenoy, Peaktk: An open source toolkit for peak forecasting in energy systems, in: ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS), 2022, pp. 324–339.
[226] d. O. Erick Meira, C. O. Fernando Luiz, Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods, Energy 144 (2018) 776–788.
[227] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, S. Ahrentzen, Random forest based hourly building energy prediction, Energy and Buildings 171 (2018) 11–25.
[228] F. Z. Abera, V. Khedkar, Machine learning approach electric appliance consumption and peak demand forecasting of residential customers using smart meter data, Wireless Personal Communications 111 (2020) 65–82.
[229] J. R. Vergara, P. A. Estévez, A review of feature selection methods based on mutual information, Neural computing and applications 24 (2014) 175–186.
[230] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on pattern analysis and machine intelligence 27 (2005) 1226–1238.
[231] A. Y. Ng, Feature selection, l 1 vs. l 2 regularization, and rotational invariance, in: Proceedings of the twenty-first international conference on Machine learning, 2004, p. 78.
[232] L. Hu, Z. Zhen, K. Li, F. Wang, An ultra-short-term pv power prediction model based on path space distance cross-similar clustering and stl decomposition, in: 2019 IEEE Sustainable Power and Energy Conference (iSPEC), IEEE, 2019, pp. 1353–1358.
[233] J. Mejia, L. Avelar-Sosa, B. Mederos, E. S. Ramírez, J. D. D. Roman, Prediction of time series using an analysis filter bank of lstm units, Computers & Industrial Engineering 157 (2021) 107371.
[234] B. Ibrahim, L. Rabelo, A deep learning approach for peak load forecasting: A case study on panama, Energies 14 (2021) 3039.
[235] A. Aziz, D. Mahmood, M. S. Qureshi, M. B. Qureshi, K. Kim, Ai-based peak power demand forecasting model focusing on economic and climate features, Frontiers in Energy Research 12 (2024) 1328891.
[236] A. Bellahsen, H. Dagdougui, Aggregated short-term load forecasting for heterogeneous buildings using machine learning with peak estimation, Energy and buildings 237 (2021) 110742.
[237] H. J. Azemena, A. Ayadi, A. Samet, Explainable artificial intelligent as a solution approach to the duck curve problem, Procedia Computer Science 207 (2022) 2747–2756.
[238] Y. Amara-Ouali, M. Fasiolo, Y. Goude, H. Yan, Daily peak electrical load forecasting with a multi-resolution approach, International Journal of Forecasting 39 (2023) 1272–1286.
[239] L. Cai, J. Gu, Z. Jin, Two-layer transfer-learning-based architecture for short-term load forecasting, IEEE Transactions on Industrial Informatics 16 (2019) 1722–1732.
[240] C. Li, G. Li, K. Wang, B. Han, A multi-energy load forecasting method based on parallel architecture cnn-gru and transfer learning for data deficient integrated energy systems, Energy 259 (2022) 124967.
[241] T. Miller, I. Durlik, A. Łobodzińska, L. Dorobczyński, R. Jasionowski, Ai in context: harnessing domain knowledge for smarter machine learning, Applied Sciences 14 (2024) 11612.
[242] Y. Bai, S. Camal, A. Michiorri, News and load: A quantitative exploration of natural language processing applications for forecasting day-ahead electricity system demand, IEEE Transactions on Power Systems 39 (2024) 6222–6234.
[243] D. Wang, J. Gan, J. Mao, F. Chen, L. Yu, Forecasting power demand in china with a cnn-lstm model including multimodal information, Energy 263 (2023) 126012.
[244] K. Mayer, L. Haas, T. Huang, J. Bernabé-Moreno, R. Rajagopal, M. Fischer, Estimating building energy efficiency from street view imagery, aerial imagery, and land surface temperature data, Applied Energy 333 (2023) 120542.
[245] S. Dai, M. Eames, R. Vinai, V. I. Sucala, Energynet: A modality-aware attention fusion network for building energy efficiency classification, Applied Energy 379 (2025) 124888.
[246] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
[247] R. Agarwal, L. Melnick, N. Frosst, X. Zhang, B. Lengerich, R. Caruana, G. E. Hinton, Neural additive models: Interpretable machine learning with neural nets, Advances in neural information processing systems 34 (2021) 4699–4711.
[248] E. S. Ali, M. Elkholy, T. Senjyu, S. M. A. Elazim, E. S. Hassan, Z. Alaas, M. E. Lotfy, A flexible multi-agent system for managing demand and variability in hybrid energy systems for rural communities, Scientific Reports 15 (2025) 16255.
[249] M. Franceschelli, A. Pilloni, A. Gasparri, Multi-agent coordination of thermostatically controlled loads by smart power sockets for electric demand side management, IEEE Transactions on Control Systems Technology 29 (2020) 731–743.
[250] J. Lim, I. Doh, K. Chae, Security system architecture for data integrity based on a virtual smart meter overlay in a smart grid system, Soft Computing 20 (2016) 1829–1840.
[251] D. Banisar, S. Davies, Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments, J. Marshall J. Computer & Info. L. 18 (1999) 1.
[252] F. Hu, M. Qiu, J. Li, T. Grant, D. Taylor, S. Mccaleb, L. Butler, R. Hamner, A review on cloud computing: Design challenges in architecture and security, Journal of Computing & Information Technology 19 (2011) 25–55.
[253] Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: Concept and applications, ACM Transactions on Intelligent Systems and Technology (TIST) 10 (2019) 1–19.
[254] S. Dai, F. Meng, Q. Wang, X. Chen, Federatednilm: A distributed and privacy-preserving framework for non-intrusive load monitoring based on federated deep learning, in: 2023 International Joint Conference on Neural Networks (IJCNN), IEEE, 2023, pp. 01–08.
[255] S. Dai, F. Meng, Addressing modern and practical challenges in machine learning: a survey of online federated and transfer learning, Applied Intelligence (2022) 1–28.
[256] D. Wu, W. Lin, Efficient residential electric load forecasting via transfer learning and graph neural networks, IEEE Transactions on Smart Grid 14 (2022) 2423–2431.
[257] W. Lin, D. Wu, M. Jenkin, Electric load forecasting for individual households via spatial-temporal knowledge distillation, IEEE Transactions on Power Systems 40 (2024) 572–584.