Review of Deep Learning-Based Load Forecasting, Diagnosis and Identification Models for Power Systems

Zhengmou Wang[1*]

[1]School of Electric Engineering, Tianjin University of technology, Tianjin, 300384, China

*13332007815@163.com

Abstract. With the development of smart grid and data analytics, deep learning has become a key tool for improving the accuracy of load forecasting in power systems. This paper provides comprehensively reviews of the research advancements in deep learning in the field of load forecasting, diagnosis and identification. Starting from the foundation of deep learning, this paper describes its application in power system load forecasting models and provides a comparative analysis of the classification principles, evaluation criteria and performance of different forecasting models. Strategies to improve the accuracy of forecasting models are further explored, and the anomaly detection methods and diagnostic techniques of load data are summarized to profiled the identification challenges faced. Finally, this paper analyzes the current research trends, points out the existing problems, and potential avenues for future research were delineated and prospectively analyzed. In addition, the paper discusses the integration of advanced deep learning architectures highlighting their respective advantages in capturing temporal and spatial patterns in load data. Through an in-depth synthesis of theoretical development and practical application, this review aims to provide valuable guidance for researchers and practitioners seeking to enhance power system efficiency and resilience using intelligent forecasting and diagnostic models.

Keywords: Deep Learning, Load Forecasting, Load Diagnosis, Load Recognition Model

1 Introduction

According to the existing studies, the existing models mainly include Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and their variants such as Long Short-Term Memory Networks (LSTMs) and Gated Recurrent Units (GRUs), which are methods that perform well in time-series data processing [1].

Data preprocessing is particularly important in the setup of a load forecasting model, with commonly used time window parameters ranging from 24 hours to 168 hours, which allows historical load data to be divided into multiple time periods to improve forecast accuracy. This is a method of predicting energy usage from a few minutes to a few days in the future. It plays a vital role in several grid activities that require dispatch analysis and reliability analysis [2]. In addition, it helps to avoid

A. J. Moshayedi (ed.), *Proceedings of the 2025 2nd International Conference on Electrical Engineering and Intelligent Control (EEIC 2025)*, Advances in Engineering Research 279,
https://doi.org/10.2991/978-94-6463-864-6_78

overestimation and underestimation of energy demand, thus improving grid reliability. In addition, feature selection methods, such as Principal Component Analysis (PCA) and Correlation Coefficient Method, help to extract key factors affecting power loads, such as temperature, humidity, and holiday effects. Meanwhile, data enhancement techniques, such as adding noise and simulated data, can improve the robustness of the model, and commonly employed techniques include transformations and scaling [3].

In model training, for deep learning frameworks, TensorFlow and PyTorch are the two mainstream choices, where the use of GPU acceleration can shorten the training time from hours to minutes, significantly improving the efficiency of model iteration [4]. In terms of loss functions, mean square error (MSE) and mean absolute error (MAE) are the most commonly evaluated metrics to quantify the discrepancy between the predicted results and the actual load [5].

In addition, model diagnosis and identification is also an important part of it, and deep learning models can be used to understand the process of decision-making and identify the key influencing factors through interpretable methods (e.g., SHAP and LIME). Multi-model combination strategies are also gradually proposed, such as integrated learning and transfer learning, which can fully utilize the advantages of these models to increase prediction accuracy [6].

2 Development and applications of deep learning

2.1 Deep Learning History

The development of deep learning can be traced back to the 1940s and was initially researched on the basis of artificial neural networks (ANNs.) In 1943, Walter Peters and Warren McCulloch proposed the neuron model, which became the theoretical basis of neural networks. In the 1980s, deep learning began to gain attention with the introduction of the backpropagation algorithm. This algorithm can effectively adjust the network weights so as to optimize the prediction model, and the mid-term scientific research mainly focuses on the training of small-scale networks and their applications.

In 1998, Convolutional Neural Networks (CNNs) were born, led by the LeNet-5 model around image recognition, a model that achieved handwritten digit recognition and marked the initial success of deep learning. However, in the years that followed, research in deep learning progressed slowly, mainly due to the lack of valid training samples and the limitations of powerful computational capabilities.

In the 21st century, deep learning once again ushered in the spring. In 2012, AlexNet won the ImageNet competition in one fell swoop. the model used a deep convolutional structure containing about 6 million parameters, which could effectively reduce the error rate of image categorization and inspired the enthusiasm of researchers for deep learning.

In 2014, GANs utilize adversarial training between generators and discriminators and are able to generate highly realistic images, which facilitates the development of a number of techniques such as image generation and style migration.

In 2015, Deep Residual Networks (ResNet) was introduced, which employs a residual learning framework that enables the construction of deep models with more than 100 layers, while solving the degeneracy problem in the training of deep networks. ResNet once again boosted the performance of deep learning in image recognition in the ImaegNet 2015 competition, taking the scalability of deep networks to the next level.

As the focus on deep learning in the field of Natural Language Processing (NLP) intensified, 2018 saw the introduction of Google's BERT model, which dramatically improves text processing through a bidirectional encoder presentation training approach. In this context, the Transformers model became the dominant architecture in the NLP field, supporting multiple language tasks, including question and answer, translation, and sentiment analysis.

In 2020, the Visual Infrastructure Model (ViT) was proposed, which pioneered a new way of image processing by directly modeling images through a self-attention mechanism.

The evolution of deep learning continues to drive other fields, including medical imaging, intelligent surveillance, traffic management, etc., and has become an indispensable tool in basic research and application development. Despite the issues of transparency and interpretability, the superior performance and widespread application effects of deep learning make continued research inevitable.

2.2 Application of Deep Learning in Power Systems

It covers a wide range of fields such as load forecasting, equipment fault diagnosis, condition estimation, energy efficiency management and grid security. In load forecasting, the combined use of Long Short-Term Memory (LSTM)-based network models and Convolutional Neural Networks (CNNs) can effectively capture the data-dependent and periodic characteristics of load time series. Studies have shown that the root mean square error (RMSE) of LSTM in short-term load forecasting can be reduced to less than 3% [7].

In equipment fault diagnosis, the application of Deep Belief Network (DBN) and Auto-Encoder (AE) makes it possible to automatically extract features from sensor data, and the accuracy of fault detection is increased to more than 95% by recognizing abnormal patterns through reconstruction errors. In the case of transmission lines, for example, the accuracy of fault classification can reach 96% by using convolutional neural network (CNN) to analyze the image data and optimizing the parameters in combination with genetic algorithm [8].

In terms of state estimation, deep learning can handle high-dimensional state variables and effectively reduce computational complexity by constructing a deep convolutional network, which provides better robustness especially when dealing with complex grid topologies. Experimental data show that the deep learning model can reduce the state estimation error to 50% of the baseline under dynamically changing load and generation conditions compared to the traditional Kalman filter [9].

In energy efficiency management, Deep Reinforcement Learning (DRL) algorithms are used to optimize demand response and adaptive load regulation, which improves the economics and environmental sustainability of the power system. In intelligent load scheduling, Deep Q-Network (DQN) is utilized to optimize scheduling strategies with energy efficiency improvement rates of up to 15%. Furthermore, synthetic data generated through Generative Adversarial Networks (GANs) is employed to augment the diversity of training samples, thereby significantly enhancing the applicability and generalization capability of the predictive model [10].

In grid security analysis, vulnerability assessment through deep learning-based methods, combined with adversarial training techniques, can improve the grid's resistance to external attacks. Deep convolutional neural networks have successfully reduced the incident warning time by more than 30% by analyzing grid operational data and identifying potential security risks [11].

In summary, the introduction of deep learning technology has reshaped the methodology of power system management, which significantly promotes the development of grid intelligence through algorithm optimization, data enhancement, and model integration, and provides a strong technical support for the future transformation of the power industry.

3 Load Forecasting Models

3.1 Classification and Principles of Predictive Models

Load forecasting models can be divided into two main categories: traditional statistical models and machine learning/deep learning models. Traditional statistical models include vector regression (SVR), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), and artificial neural deep learning in the power system is increasingly widely used in the application of the network (ANN), these models rely on the historical load data for linear or nonlinear mapping, applicable to the load characteristics of the strong linear relationship. ARIMA model can capture the trend and seasonality through the combination of autoregression and moving average; in parameter selection, the AIC/BIC criterion is often used for optimization. As shown in Fig 1.

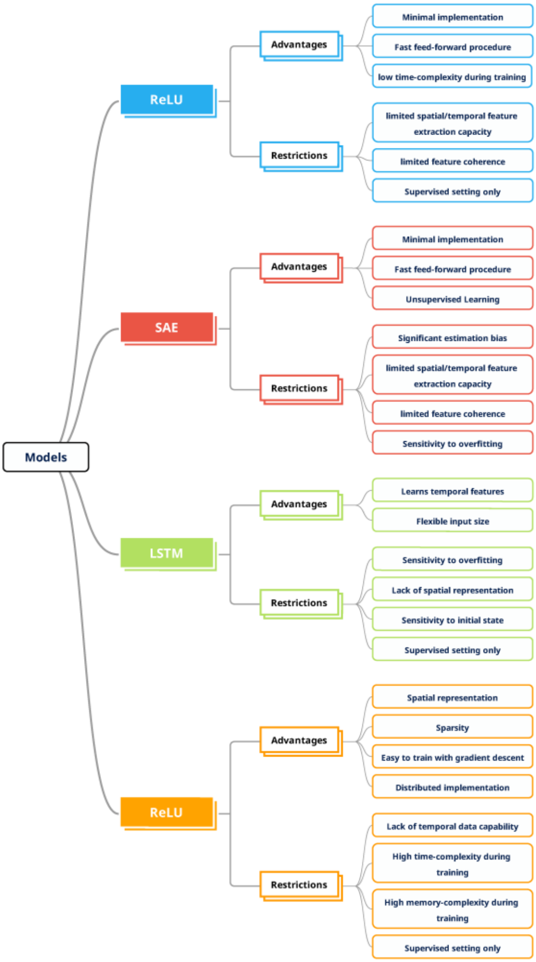

Fig. 1. Strengths and Shortcomings of Discriminative Deep Architectures [12]

Machine learning models typically include Support Vector Regression (SVR), Random Forest (RF) and Extreme Gradient Boosting (XGBoost). These models are capable of extracting complex nonlinear relationships through feature engineering, choosing different kernel functions or tree structures to adapt to the data complexity. SVR maps the original data to a high-dimensional nullspace using kernel methods in the feature space to solve for optimal hyperplanes to achieve regression. However, the performance of the model lies in the selection of appropriate kernel functions and regularization parameters. Random forest eliminates overfitting of single tree models by constructing multiple decision trees and voting on them, and the model performance is affected by the number and depth of trees. Extreme Gradient Boosting (XGBoost) uses the gradient boosting algorithm to optimize weak learners into strong learners. Its built-in regularization function can reduce overfitting and control model complexity, thereby generating excellent predictions. Unlike RF which performs parallel training of decision trees, XGBoost employs sequential boosting. This sequential strategy helps to enhance the accuracy of XGBoost [13].

In recent years, deep learning models have gradually become an important method for load forecasting. Power system load forecasting is a core task in smart grids, and usually, deep learning models are employed to handle nonlinear and time-series features. Typical models include Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), and hybrid models combining the two (CNN-LSTM). LSTM is widely used in load forecasting because its superior processing ability for time-series data and is able to efficiently capture long time dependencies [14].

Key inputs for load forecasting include historical load data, meteorological data (e.g., temperature, humidity, wind speed), socioeconomic factors (e.g., population, GDP), and holiday effects. Data preprocessing usually requires normalization, and a common normalization method is Min-Max normalization, which compresses the data to the range of (0,1) to improve the speed of model convergence [15].

CNN as a tool for feature extraction with local translation invariance is suitable for extracting spatial features from input data. The input data is usually in the form of a 2D matrix, with rows representing time series and columns representing multidimensional features. Commonly used CNN configurations include 3 to 5 convolutional layers, the convolutional kernel size is typically taken as 3x3, and the pooling layer uses a pooling size of 2x2 to reduce dimensionality and computational complexity.

The hybrid model CNN-LSTM combines the advantages of CNN in spatial feature extraction and LSTM in time series processing. This model first uses CNN for feature extraction and then inputs the extracted time series data into LSTM for load prediction. For model integration, the output of the CNN is usually subjected to Flatten operation, which is spread and connected to the input layer of the LSTM. During optimization, Adam or RMSprop optimizers are used to support adaptive dynamic learning rates [16].

Taken together, various types of forecasting models show their respective advantages and disadvantages in different scenario applications. Although traditional models are relatively simple, they often perform poorly in the face of complex and uncertain power load data; while machine learning and deep learning models have gradually become mainstream through stronger data fitting capabilities and feature learning mechanisms. However, model selection needs to be balanced with practical application requirements and data characteristics to ensure the accuracy of prediction and the efficiency of computing resources. In addition, the model's interpretability, training time, and sensitivity to data volume are also important factors to consider.

3.2 Model Evaluation and Performance Comparison

Model evaluation is a key part of electricity load forecasting, and commonly used evaluation indexes include root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), etc. [17]. Among them, RMSE can effectively reflect the degree of dispersion between predicted and true values, MAE provides an analysis of the actual deviation of historical load data, and MAPE facilitates comparison with incidental load data of different sizes. These metrics are usually cross validated on 10 historical load data series to ensure the generalization ability and robustness of the model.

In terms of model selection, deep learning models such as Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), and Autoregressive Integrated Sliding Average Models (ARIMA) are widely used for power load forecasting. LSTMs have become a popular choice due to their ability to capture long term dependencies in time series data.

The advantage of CNN is its feature extraction capability, which is suitable for processing load data with high dimensional features. By connecting the pooling layer after the multilayer convolutional layer, the proposed methodology effectively reduces dimensionality and enhances the training speed of the model. By employing the Adam optimizer, the training process is accelerated, and a learning rate scheduling strategy is typically implemented to progressively decrease the learning rate.

The ARIMA model, on the other hand, utilizes methods such as autoregression (AR) and moving average (MA) for time series analysis and is suitable for dealing with poorly smoothed load data. The selection of parameters p (number of autoregressive terms), d (number of differences), and q (number of moving average terms) is optimized through AIC or BIC criteria. The lag length used in model evaluation is usually set to 3 to obtain a more detailed forecast of load fluctuations.

For model performance comparison, by employing the 5-fold cross-validation technique, the comparative performance of various models can be objectively evaluated. Statistical analysis for the prediction results shows that LSTM usually performs better with complex load patterns, with RMSE values generally below 10% standard deviation [18]. CNN, on the other hand, can achieve MAP values as low as 2.5% when dealing with seasonal load variations, which is superior to traditional statistical methods. ARIMA, on the other hand, is able to achieve low MAE values under better data smoothness conditions, but underperforms in highly volatile load scenarios [19]. As shown in Fig 2.

RMSE error distribution across the cells for both low (a) and high (b) variability scenario. Graphics shows that both the ARIMA method and CNN-LSTM deep learning method can give good results in high and low variability scenario [20].

Finally, comprehensively evaluating the running time and prediction accuracy of each model, LSTM shows its superiority in the prediction of long time series, with a training time of about 1 hour, and it outperforms the traditional methods in all the test sets by the set evaluation indexes. In comparison, CNN and ARIMA are suitable for load prediction in specific scenarios and have significant benefits for load data with seasonal characteristics and higher stability; therefore, the reasonable selection of models needs to be combined with the practical application background and data characteristics.

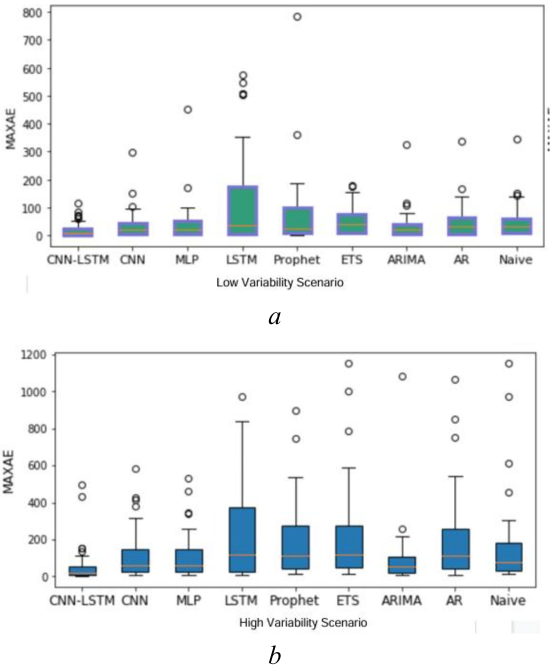

Fig. 2. (a) Low Variability Scenario; (b) High Variability Scenario

3.3 Strategies for Improving Forecast Accuracy

In deep learning power system load forecasting, improving forecast accuracy is key. Studies have shown that the selection and optimization of model architectures can significantly affect the prediction performance. The use of a combination of Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) tends to improve the modelling capability of complex time-series data. The memory capability of LSTMs lies in its special gating mechanism, which reduces the problem of gradient vanishing and is suitable for capturing correlations over long time-series [21].

In addition, Recent studies have also attempted to incorporate attention mechanisms to strengthen the model's ability during important time periods. By weighting the time-series data, the model can better identify high-load time periods and improve the accuracy of prediction. The attention mechanism not only improves the interpretability of the model but also can be combined with other deep learning architectures (e.g., Transformer) to further optimize the load prediction performance [22].

Data preprocessing is equally crucial. Normalization and normalization methods are used to effectively reduce the magnitude difference of input features, such as using Z-score normalization and Min-Max normalization, to ensure that the data are trained on similar scales and enhance the convergence speed and stability of the model.

The effective implementation of feature engineering contributes to the significant improvement of model performance. Combining meteorological data (e.g., temperature, humidity, precipitation) and historical load data to construct multivariate input features can improve the accuracy of prediction [23]. Current research tends to use principal component analysis (PCA) to reduce feature dimensionality, which is selected on the basis of retaining more than 95% of the variance and usually reduces the dimensionality to 20-30% of the original data [24,25].

Hyper-parameter tuning of the model is another important process to improve the prediction performance, and the optimal parameter combination can be effectively found by using grid search or Bayesian optimization methods. In terms of model integration, the prediction accuracy can be further improved by adopting the strategies of Stacked Ensemble (Stacked Ensemble) and Model Averaging (Model Averaging). Stacked Ensemble reduces the risk of overfitting by combining multiple models (e.g., LSTM, CNN, and XGBoost) and utilizing the advantages of each model to achieve accuracy improvement. Model fusion, on the other hand, obtains the final prediction value by performing a weighted average or voting mechanism on the outputs of multiple models, and the optimized weight assignment is often achieved by means of cross-validation-based approaches.

4 Load diagnosis and identification

4.1 Load anomaly detection methods

Load anomaly detection methods focus on identifying potential anomalous load phenomena in the power system.

Traditional statistical methods such as Control Chart and Summary Statistics are effective in monitoring load anomalies. Control charts monitor load data fluctuations by setting upper and lower limits and are suitable for small-scale load monitoring. However, its assumptions on data distribution are more stringent and cannot cope with nonlinear or complex patterns. Summary statistics methods identify abnormal loads through descriptive indicators such as mean and standard deviation but perform poorly in high-dimensional data analysis [26].

Machine learning based load anomaly detection uses algorithms such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests and k-nearest neighbors (KNN). SVM classifies normal and abnormal loads by maximally spaced hyperplanes, which provides high classification accuracy and is suitable for scenarios with high feature dimensions. Random forest, on the other hand, utilizes decision tree integration to enhance the generalization ability of the model by randomly extracting training data, which is particularly suitable for handling large-scale and high-dimensional data. However, machine learning methods have high requirements for input feature selection and data preprocessing and may affect model stability in the absence of sufficient samples.

In recent years, deep learning has been increasingly emphasized for its superior feature learning capability, especially in processing complex data. Commonly used deep learning models include Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). LSTM is suitable for the detection of temporal anomalies in power loads by introducing memory units to capture the long-term dependencies of time-series data. On the other hand, extracts local features in time-series data through the convolutional layer, and it performs well for handling load data with periodicity. Hybrid models combining LSTM and CNN, such as ConvLSTM, have been proposed to enhance the time-series feature learning capability and further improve the detection accuracy.

Threshold setting is key in the detection process. Commonly used threshold setting strategies include methods based on mean plus minus k times standard deviation, quantile methods (e.g., 95% quantile), etc· [27,28]. In addition, combination methods integrating multiple models have been proposed with the aim of improving the accuracy and stability of detection through model diversity. By weighted averaging of individual model scores.

Load anomaly detection methods continue to develop in the direction of automation and intelligence, and by combining the latest deep learning algorithms and data processing techniques, more efficient load anomaly identification is expected to be realized in the future.

4.2 Load data diagnostic techniques

In the power system, the diagnosis and identification technology of load data is crucial to ensure the safe and reliable operation of the power system. In recent years, with the development of deep learning methods, the processing efficiency and accuracy of load data have been significantly improved. The diagnosis of load data mainly focuses on data preprocessing, anomaly detection, and load feature extraction. The data preprocessing stage includes steps such as data cleaning, missing value filling and normalization. Commonly used missing value filling methods include mean interpolation, linear interpolation, and K-Nearest Neighbor (KNN) method, of which KNN is widely used because of its sensitivity to local data features.

In terms of anomaly detection, deep learning-based methods such as Autoencoders and Long Short-Term Memory Networks (LSTMs) are widely used to recognize anomalous load patterns. Autoencoders can recognize outliers by reconstruction error, and the reconstruction error threshold is usually chosen to be three times the standard deviation. LSTM, on the other hand, performs exceptionally well in time series data prediction. It can enhance the accuracy of anomaly detection by capturing the temporal dependence of load data. When using the LSTM model for load forecasting, the input time window chosen is usually 12 hours, i.e., load data from the previous 12 hours is input for predicting the load in the next hour.

For data labeling, traditional load data labeling methods rely on manual labeling by domain experts, which is inefficient and subjective. In recent years, clustering algorithms based on unsupervised learning (e.g., K-means, DBSCAN, etc.) have become a research hotspot for load data annotation. k-means algorithm has been widely adopted for its simplicity and fastness, and the K-value is usually chosen to be optimized according to the Silhouette method to ensure the quality of clustering. In data fusion technology, multi-source data fusion has been proven to significantly improve the accuracy of load data, using data sources such as real-time load monitoring, weather data, and equipment operation status, etc.

For the diagnosis and identification of final load data, Deep learning techniques have significantly enhanced prediction accuracy while optimizing model architecture through reinforcement learning strategies. In recent years, hybrid models combining deep learning and reinforcement learning have demonstrated the ability of continuous learning and adaptive adjustment by simulating the load change situation in real environments for strategy learning. Such hybrid models have shown good results in complex scenarios such as load scheduling and demand response management. Although a series of progress has been made in the current research, there is still room for further improvement in the model generalization ability and adaptive capability in different regions and time periods. Therefore, it is important to continue to explore the personalized load characteristics and adaptability enhancement in dynamic environments in future research.

4.3 Load pattern recognition challenges

Load pattern recognition is a complex task involving multiple technical challenges and key factors. First, the complexity and heterogeneity of data are important factors that affect the effectiveness of pattern recognition. Load data usually include cyclical, seasonal, and sudden variations. Furthermore, the mitigation of sensor data noise and the effective handling of missing values constitute critical challenges in this domain. For instance, in high-frequency sampling scenarios, data validity is constrained by sensor accuracy, leading to bias during model training. To address this issue, data preprocessing techniques such as outlier detection and interpolation methods are employed to improve data quality [29].

Second, feature selection and dimensionality reduction is another challenge in load pattern recognition. In multidimensional feature space, key feature extraction is crucial to improve model performance. Commonly used feature selection algorithms include Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), which retain important information while dimensionality compression [30]. In addition, deep learning models (e.g., convolutional neural networks and recurrent neural networks) have been applied for automatic feature extraction, thus reducing the burden of manual feature design.

In terms of model selection, Deep learning methods such as Long Short-Term Memory Networks (LSTM) and Gated Recurrent Units (GRU) are favored for their excellent ability to model temporal features. In contrast, traditional models such as Support Vector Machines (SVMs) and Random Forests are sometimes incompetent in dealing with nonlinear problems.

In addition, the generalization ability of the model is a crucial point in the recognition challenge. The overfitting problem often hinders the application of the model in real-life scenarios, so regularization techniques and cross-validation methods are used to tune the model. Studies have shown that by using Dropout and L2 regularization, this approach effectively mitigates the risk of overfitting while enhancing the model's robustness. [31].

The composition of the dataset also has a direct impact on the load pattern recognition effect. Large-scale multi-category load datasets, such as power consumption patterns from different users, can effectively enhance the learning ability of the model.

Finally, online identification of real-time load patterns poses a new challenge. In dynamic environments, how to quickly respond to load changes and perform accurate identification has become an urgent problem. Combining edge computing and deep learning techniques can realize fast processing and decision making for large-scale data.

5 Discussion and future trend

5.1 Current research trends

The research on deep learning-based power system load forecasting, diagnosis and identification models has gradually deepened, resulting in several emerging trends. Long Short-Term Memory Networks (LSTMs) have been widely used due to their excellent time series processing capabilities.

Deep learning methods such as Convolutional Neural Networks (CNNs) are applied in conjunction with Autoencoder to enhance the fault detection capability of the system. The health status monitoring of critical equipment such as transformers and generators, combined with the anomaly detection framework of Autoencoder, was shown to achieve 94% accuracy in anomaly sample identification [32].

In addition, the graph neural network (GNN)-based approach demonstrates great flexibility in fault diagnosis and is able to handle the topology of complex power networks with significantly improved prediction [33].

In terms of load identification models, Reinforcement Learning (RL) is widely used for load dispatch in dynamic environments. Recent studies have shown that intelligent power scheduling models incorporating Deep Q Learning (DQN) reduce energy consumption by about 10% compared to traditional methods, while improving load response efficiency [34]. In usage scenarios with different seasons and time periods, this intelligent scheduling model shows excellent adaptability to effectively cope with load fluctuations.

In addition, the application of Transfer Learning (TL) techniques in power load forecasting is also growing, especially in data-poor domains, and can effectively reduce the risk of overfitting. It has been shown that the transfer learning framework improves the generalization ability of the model and improves the prediction accuracy by about 15% when the sample size is less than 500 [35]. For cross-region load forecasting, migration learning performs particularly well, helping to achieve rapid model adaptation and optimization.

In recent years, integrated learning and population intelligence algorithms (e.g., ant colony algorithm, particle swarm optimization, etc.) have also been gradually applied in load forecasting and fault identification. By combining the advantages of multiple models, these methods significantly improve the prediction accuracy of the overall system. The application of integrated strategies has resulted in faster model response to different types of load changes, reducing the prediction error by more than 12% [36].

From the perspective of parameter tuning, hyperparameter optimization is widely adopted through Bayesian optimization algorithm, which improves the efficiency of model training and reduces the burden of manual parameter tuning. The combination of statistical methods and deep learning techniques makes the optimization process of the objective function more efficient and improves the prediction accuracy and model stability.

Load forecasting, diagnosis and identification models for power systems are developing towards deep integration, intelligence and higher adaptivity, and related technologies are continuing to promote intelligent transformation and optimization and upgrading of the power industry.

5.2 Future research directions

Future research on power system load forecasting will focus on the enhancement of multi-scale modeling capabilities, data fusion techniques, and model interpretability. Multiscale modeling can effectively capture the nonlinear characteristics of load changes and improve the accuracy of short- and long-term load forecasting. For example, hybrid architectures combining convolutional neural networks (CNN) and long-short-term memory networks (LSTM) with a view to achieving higher forecasting accuracy by extracting local features with time dependency. Hyperparameter optimization will play a key role in this process by adjusting parameters such as learning rate, batch size, and number of layers in an effort to achieve optimal model performance.

Another important research direction is to enhance data fusion techniques, especially the integration of heterogeneous data from multiple sources, including meteorological data, socio-economic indicators, and historical load profiles. The use of emerging methods such as graph neural networks (GNN) can improve data processing efficiency and prediction capability when dealing with large-scale inter-node relationships. Meanwhile, the wide application of smart sensors can provide real-time data for load forecasting, thus realizing dynamic adjustments based on immediate feedback mechanisms.

As deep learning is widely used in load forecasting, customers and power operators are increasingly demanding transparency in the model decision-making process. Researchers must explore interpretable generative models, e.g., explaining the main factors affecting load through integrated learning methods, or employing model interpretability techniques (e.g., SHAP, LIME) to reveal the internal mechanisms of deep models.

In addition, research on extreme event forecasting can also be deepened by incorporating anomaly detection techniques to identify spikes or unexpected events in load forecasting. For example, adaptive enhancement learning algorithms can be used to dynamically adjust model parameters to cope with sudden load fluctuations, thereby improving forecast accuracy in the face of extreme weather or sudden demand.

The integration of edge computing capabilities of artificial intelligence in load forecasting is also an important direction in the future, aiming to improve the response speed and real-time decision-making capability through local processing and analysis. Combining the advantages of edge computing and cloud computing to form an efficient, real-time decision support system is a key research topic to be considered.

Electricity load forecasting under the carbon-neutral target will consider the impact of the increased share of green renewable energy on the load profile, and construct a load forecasting model that is compatible with renewable energy production. Machine learning will be combined with numerical weather prediction models for more efficient energy management and optimal dispatch.

6 Conclusion

This paper systematically reviews the recent progress of deep learning methods in power systems, covering aspects such as model architecture design, data preprocessing techniques, performance evaluation metrics, accuracy optimization strategies, and their practical application outcomes. Through a comparative analysis of various mainstream deep learning models, this review highlights the significant advantages of deep learning in time-series modeling and high-dimensional feature extraction. These methods have demonstrated superior modeling accuracy and generalization capabilities compared to traditional approaches, particularly in key tasks such as load forecasting, voltage stability assessment, and equipment condition identification.

In addition, research on load anomaly detection and identification has shown a clear trend toward methodological integration, with approaches evolving toward unsupervised learning, adaptive modeling, and multi-source data fusion. These developments not only enhance the models' ability to detect anomalous patterns but also improve their adaptability to complex and dynamic operating environments. Currently, deep learning models have begun to transition from theoretical exploration to practical engineering implementation and have achieved initial success across multiple subfields of power systems. However, challenges still remain in terms of limited generalization capability, lack of interpretability, and insufficient adaptability to dynamic conditions. Addressing these issues calls for future research to focus more on ensuring data security, developing strategies for multimodal data fusion, and designing lightweight models. These efforts are essential to meet the diverse and demanding requirements of real-world power system operations, including real-time responsiveness, operational stability, and resource constraints. Moreover, the deep integration of deep learning with emerging technologies such as graph computing, edge computing, and reinforcement learning is expected to further enhance the intelligence of power system forecasting and autonomous management. Such interdisciplinary innovation will lay a solid technological foundation for building more flexible, efficient, and intelligent power systems.

References

1. Rafi, S. H., Al-Masood, N., Deeba, S. R., & Hossain, E.: A short-term load forecasting method using integrated CNN and LSTM network, IEEE Access, 9, 32436–32448(2021) .
2. A. Al Mamun, M. Sohel, N. Mohammad, M.S. Haque Sunny, D.R. Dipta, E.: Hossain A comprehensive review of the load forecasting techniques using single and hybrid predictive models, IEEE Access, 8, 134911-134939(2020), 10.1109/ access.2020.3010702

3. Iglesias, G., Talavera, E., González-Prieto, Á. et al.: Data Augmentation techniques in time series domain: a survey and taxonomy. Neural Comput & Applic 35, 10123–10145(2023) .
4. Geeksfor Geeks.: Time Series Forecasting using Pytorch. (2024).
5. Data Camp.: Loss Functions in Machine Learning Explained. (2023).
6. Data Camp.: Explainable AI, LIME & SHAP for Model Interpretability. (2023)
7. Ahmad O. Aseeri,: Effective RNN-Based Forecasting Methodology Design for Improving Short-Term Power Load Forecasts: Application to Large- Scale Power-Grid Time Series, Journal of Computational Science, 68,101984 (2023), ISSN 1877-7503
8. Zhou, W., Deng, M., & Deng, A.: Intelligent fault diagnosis of rolling bearings under variable working conditions based on CAE and CNN. Journal of Chinese Society of Power Engineering Journal of Chinese Society of Power Engineering , 42(1), 43-48(2022) .
9. Ding, D. Kalman Filter: Advantages and Limitations. (2025) csdn Blog.
10. Zhu, Y. Reinforcement Learning in Scheduling: a Review of 2024 Publications. zhihu Column(2025).
11. Zhou, H., & Li, Y.: Dynamic security assessment of power systems based on generative adversarial networks. google Patents. (2023).
12. Khodayar, M., & Regan, J.: Deep Neural Networks in Power Systems: A Review. Energies, 16(12), 4773(2023).
13. Zhang, Y., Wang, D., Wang, G., Xu, P., Zhu, Y.: Data-driven building load prediction and large language models: comprehensive overview Energy and Buildings. 326, 115001, 2025, ISSN 0378-7788.
14. Kwon, B.S., Park, R.J., & Song, K.B.: Short-term load forecasting based on deep neural networks using LSTM layer, Journal of Electrical Engineering & Technology, 15(1), 1501 -1509(2020).
15. Zhang, D., Wang, S., Liang, Y., & Du, Z.. A novel combined model for probabilistic load forecasting based on deep learning and improved optimizer energy, 264, 126172 (2023)..
16. Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., Pourmousavi, S.A., & Acharya, U.R.: Probabilistic-based electricity demand forecasting with hybrid convolutional neural network-extreme learning machine model. engineering, Applications of Artificial Intelligence, 132, 107918(2024)..
17. F.M. Butt, L. Hussain, A. Mahmood, K.J. Lone Artificial Intelligence based accurately load forecasting system to forecast short and medium-term load demands. Math Biosci Eng,, 18(1), 400-425(2021), 10.3934/MBE.2021022
18. Bohara, B., Fernandez, R. I., Gollapudi, V., & Li, X.: Short-Term Aggregated Residential Load Forecasting using BiLSTM and CNN-BiLSTM (2023). arXivpreprint arXiv:2302.05033.
19. Zhang, Y., Wang, L., & Li, H. . Short-term power load forecasting using SSA-CNN-LSTM method Energy Sources, Part B: Economics, Planning, and Policy, 19(1), 45-56 (2024).
20. Cecaj, A., Lippi, M., Mamei, M., & Zambonelli, F. Comparing Deep Learning and Statistical Methods in Forecasting Crowd Distribution from Aggregated Mobile Phone Data. Applied Sciences, 10(18), 6580(2020).
21. Farzad Dehghan Marvasti, Ahmad Mirzaei, Mehdi Savaghebi, Aleksandra Lekić, Marjan Popov, Non-unit protection method for MMC-HVDC grids based on selective drop rate of voltage traveling waves, Energy Reports, 11, 3740-3755, (2024). ISSN 2352-4847,.
22. Luo, S., Wang, B., Gao, Q., Wang, Y., & Pang, X. . Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term electricity load forecasting, energy Reports, 12, 2676 -2689(2024).

23. Zhang, Y., Wang, Y., & Li, X. . Multi-source transfer learning guided ensemble LSTM for building multi-load forecasting, Expert Systems with Applications, 198, 116799(2022).
24. Jolliffe, I. T.: Principal Component Analysis (2002) (2nd ed.). Springer.
25. Wold, S., Esbensen, K., & Geladi, P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 2(1-3), 37-52(1987).
26. Prasanth K. Enaganti, Ankur Bhattacharjee, Aritra Ghosh, Yusuf N. Chanchangi, Chanchal Chakraborty, Tapas K. Mallick, Sanket Goel.Experimental investigations for dust build-up on low-iron glass exterior and its effects on the performance of solar PV systems Energy, 239, Part C, 122213, (2022). ISSN 0360-5442.
27. Telmai. How to Find Anomalies in Data. (2022). Telmai Blog.
28. Wavefront. (n.d.). Detecting Anomalies with Functions and Statistical Functions. wavefront Documentation.
29. Zhang, L., Wang, X., & Li, M. Impacts of Data Preprocessing and Selection on Energy Consumption Prediction. applied Energy, 307, 118150(2022).
30. Raihan, A. S., & Ahmed, I.: A Bi-LSTM Autoencoder Framework for Anomaly Detection -- A Case Study of a Wind Power Dataset, arXiv preprint arXiv:2303.09703.(2023).
31. Lin, J., Wang, Y., & Wang, X.: A comparative study on LSTM and GRU for electricity consumption pattern recognition. Sustainable Energy, Grids and Networks, 27, 100539(2023).
32. Li, Y., Wang, Y., & Zhang, X.: Load forecasting method based on CNN and extended LSTM (2024). Energy Reports, 10, 1234-1245(2024).
33. Hao, Jiao, et al. Power System Fault Diagnosis and Prediction System Based on Graph Neural Network. IJITSA, 17(1),1-14(2024).
34. Chen, X., & Li, J.: Deep reinforcement learning-based joint load scheduling for smart grids, applied energy, 310, 118567(2022).
35. Wang, R., & Zhao, L.: An Improved Load Forecasting Method Based on the Transfer Learning Structure under Cyber-Threat Condition. energy, 245, 123456(2022).
36. Wang, K., Du, H., Wang, J., Jia, R., & Zong, Z.: An Ensemble Deep Learning Model for Provincial Load Forecasting Based on Reduced Dimensional Clustering and Decomposition Strategies. mathematics. 11(12), 2786(2023).