

Review Article

A Comprehensive Review of Deep Residual Networks for Short-term Load Forecasting

Junchen Liu¹, Faisal Arif Ahmad^{1*}, Khairulmizam Samsudin¹,
Fazirulhisyam Hashim¹, and Mohd Zainal Abidin Ab Kadir²

¹Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), 43400 UPM Serdang, Selangor, Malaysia

²Department of Electrical and Electronic Engineering, Advanced Lightning, Power and Energy Research Centre (ALPER), Faculty of Engineering, Universiti Putra Malaysia (UPM), 43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

Short-Term Load Forecasting (STLF) is a critical component of modern power systems, enabling efficient grid operation and energy management. Recent advancements in deep learning have positioned Deep Residual Network (DRN) as a promising approach for STLF, owing to their ability to capture complex and nonlinear load patterns. This paper provides a comprehensive review of DRN-based models for STLF, offering novel insights into their strengths, limitations, and future research directions. Unlike previous reviews, this work systematically evaluates DRN variants, highlighting challenges such as activation function selection, long-term dependency modelling the integration of diverse meteorological variables. Furthermore, this review proposes actionable research directions, including systematic activation function analysis, enhanced sequential modelling techniques, and multi-variable integration, to address current limitations. By bridging these gaps, this paper aims to support the development of more accurate and adaptable forecasting models, contributing to the advancement of intelligent energy management systems.

Keywords: Activation functions, DRNs, power systems, STLF

ARTICLE INFO

Article history:

Received: 04 February 2025

Accepted: 18 May 2025

Published: 06 February 2026

DOI: <https://doi.org/10.47836/pjst.34.1.14>

E-mail addresses:

gs61666@student.upm.edu.my (Junchen Liu)

faisal@upm.edu.my (Faisal Arif Ahmad)

khairulmizam@upm.edu.my (Khairulmizam Samsudin)

fazirul@upm.edu.my (Fazirulhisyam Hashim),

mzk@upm.edu.my (Mohd Zainal Abidin Ab Kadir)

* Corresponding author

INTRODUCTION

Load forecasting (LF), a key component of contemporary power systems, ensures dependable and effective grid operation. By precisely forecasting future electricity demand, LF helps power companies plan,

operate, and manage the grid, improving energy use efficiency, cutting costs, and guaranteeing supply stability. The importance and complexity of LF are growing as the demand for energy keeps rising and consumption patterns change (Ceperic et al., 2013).

The four primary forms of LF are Very Short-Term Load Forecasting (VSTLF), Short-Term Load Forecasting (STLF), Medium-Term Load Forecasting (MTLF), and Long-Term Load Forecasting (LTLF), as seen in Figure 1 and categorised based on the forecasting time range (Akhtaruzzaman et al., 2020; Liu et al., 2025). Real-time power system scheduling and operation rely heavily on STLF, which can last anywhere from one hour to a week; MTLF helps with operational planning, such as scheduling maintenance and managing the power supply to meet expected medium-term demand; MTLF typically spans one week to one year; LTLF can range from one year to many years, as it helps predict future power demands and the need for additional generating facilities; and VSTLF is used for real-time control and focusses on urgent operational demands; it can predict load up to an hour in advance. Specifically, STLF is critical to the dispatching and real-time operation of the power system. These estimates include the LF for the next seven days on a weekly basis, the LF for the next 24 hours daily, and the forecasting for several hours ahead of time (Kondaiah et al., 2022).

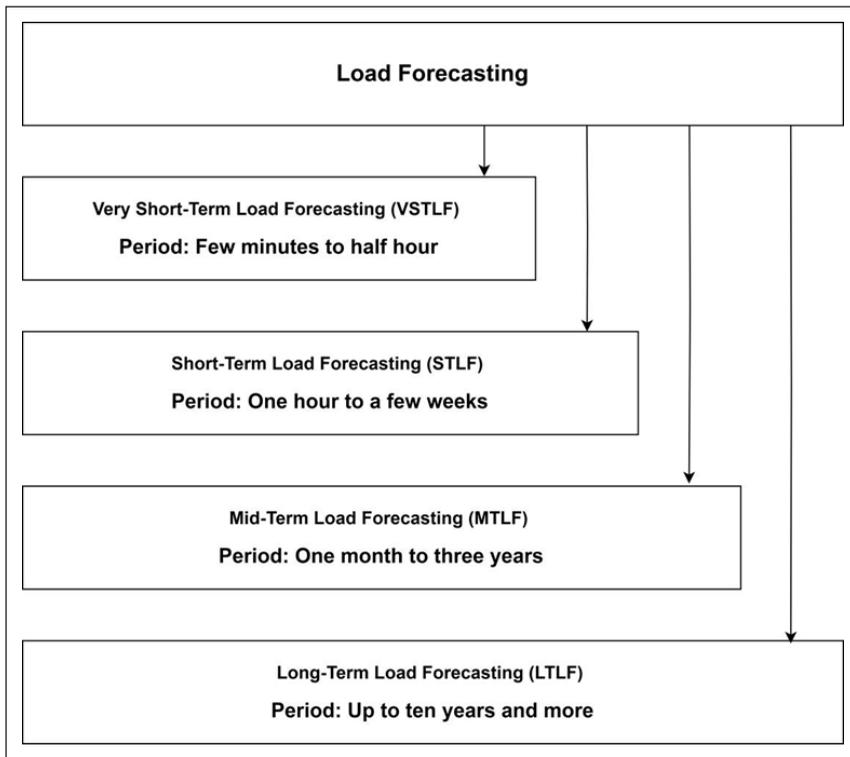


Figure 1. Classification of LF (Liu et al., 2025)

Future power system management and operation will require the ability to handle uncertainty and make decisions more quickly (ul Hassnain & Khan, 2007). Numerous application fields, including energy trading, unit commitment decisions, system security assessment, cost-effective power production distribution, and performance monitoring, depend on LF. As a result, it becomes more crucial than ever to calibrate and validate the estimated precision. Among the topics covered by LF are grid scheduling, load flow analysis, daily operations, and performance. Precise forecasting is vital to the system's normal running since imprecise LF might result in unplanned expenditures.

Traditional methods, such as linear (Song et al., 2005) or non-parametric methods (e.g., non-parametric regression (Dudek, 2012), exponential smoothing (Taylor, 2010; Taylor, 2003)), support vector regression (SVR) (Ceperic et al., 2013; Taylor, 2010; Taylor, 2003), autoregressive models (Alberg & Last, 2018), and fuzzy logic (Ali et al., 2020; Rejc & Pantos, 2011), may suffer from overfitting as the number of input variables increases, oversimplify complex load dynamics, or not generalise well (Ceperic et al., 2013; Hippert et al., 2001). STLF methodologies can be broadly divided into traditional and modern approaches.

To get around these issues, artificial neural networks (ANNs) have emerged as a potent alternative to STLF. By employing deep learning, ANN-based models effectively capture intricate load patterns, improve prediction precision, and lower overfitting risks (Hernández et al., 2014; Kuster et al., 2017). However, if a network gets more complex by adding additional inputs, nodes, or layers, overfitting problems can still occur. To improve model performance in STLF, ANN variants such radial basis function (RBF) networks (Cecati et al., 2015), wavelet-based networks (Chen et al., 2009), and extreme learning machines (ELM) (Zhao et al., 2009) have been created.

In recent years, deep neural networks (DNNs), which are characterised by their layered architecture, have gained popularity because to their ability to learn hierarchical representations of complex load data. LF research has advanced from traditional shallow designs to complicated deep learning structures that integrate several variables to represent intricate temporal and spatial relationships. This shift reflects the growing use of deep learning techniques to difficult forecasting tasks (Eren & Küçükdemiral., 2024; Fan & Hyndman, 2011; Goodfellow, 2016).

Recent advances have seen a move away from preset shallow network designs with a focuss on integrating various input into neural network topologies. Convolutional Neural Networks (CNNs), which are well-known for their capacity to extract local features, have proven to be highly effective in detecting temporal load patterns in STLF (Baul et al., 2024; Cai et al., 2019; Deng et al., 2019; Dong et al., 2017; Han et al., 2018; He, 2017; Jurado et al., 2023; Kong et al., 2020; Li et al., 2017; Shi et al., 2023; Tian et al., 2018). However, their applicability in intricate LF situations is restricted by their inadequacy to replicate long-term interactions and challenges in training deeper systems.

Recurrent Neural Networks (RNNs) (Atef et al., 2022; Bento et al., 2018; Cai et al., 2021; Ertugrul, 2016; Gan et al., 2017; Haque & Rahman, 2022; Kong et al., 2017; Kwon et al., 2020; M. Liu et al., 2023; Q. Liu et al., 2023; Marino et al., 2016; Muzaffar & Afshari, 2019; Narayan & Hipel, 2017; Tang et al., 2019; S. Wang et al., 2019; Y. Wang et al., 2019; Wu et al., 2019), such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (BiLSTM), and Bidirectional Gated Recurrent Unit (BiGRU), are excellent at capturing sequential relationships through memory cells and gating mechanisms. These designs efficiently manage both short-term and long-term dependencies in STLTF and reduce gradient vanishing issues, but they are less suitable for very lengthy sequences due to their sequential processing strategy, which raises computing needs.

With its self-attention processes, transformer models for STLTF (Cen & Lim, 2024; Fang et al., 2024; Giacomazzi et al., 2023; Huang et al., 2022; Huy et al., 2022; Jiang et al., 2022; Li et al., 2023; Liao & Radhakrishnan, 2022; Lin et al., 2024; Lu & Chen, 2024; Ran et al., 2023; Santos et al., 2023; Upadhyay et al., 2023; Valencia & Sanchez-Galan, 2022; Xu et al., 2024; Zhang et al., 2022; Zhang et al., 2023; Zhao et al., 2021) offer a robust foundation for time series forecasting by integrating long-range dependencies. Transformers work well with sequences of varying lengths, but they struggle with long sequences since the processing cost of a transformer rises quadratically with sequence length. Deeper Transformer topologies also require architectural enhancements to address stability and training problems.

CNNs, RNNs, and Transformers' scalability is often limited by the challenges presented by deeper networks, such as gradient vanishing or exploding, which limits their capacity to represent extremely complex patterns. To address these issues, Chen et al. (Chen et al., 2018) developed a unique end-to-end DRN for STLTF that incorporates residual connections. These connections effectively alleviate gradient vanishing issues and enable robust training even in deep architectures. The DRN model utilises raw historical load, temperature, and time data as inputs, reducing the need for intensive feature engineering while automatically extracting complex features from the data. However, the model's deep structure and large number of parameters introduce computational challenges, such as potential overfitting and reduced scalability. In response, researchers have proposed various DRN-based variants (Chen et al., 2018; Chen et al., 2022; Ding et al., 2021; Kondaiah & Saravanan, 2021; Kondaiah & Saravanan, 2022; Li et al., 2021; Sheng et al., 2023; Sheng et al., 2021; Tian et al., 2020; Xu et al., 2020) aimed at improving computational efficiency, enhancing scalability, and addressing overfitting, making the optimisation of DRN models an active and evolving area of research in STLTF.

This review covers the latest research on DRN in STLTF from 2018 to 2024 in the Web of Science database. The keywords used to search the papers are “short-term load forecasting” and “deep residual network”. To examine various DRN architectures and their

performance in STLF tasks, this study identifies strengths, weaknesses, and future research directions. The rest of this paper is organised as follows: Section 2 reviews different DRN models. Section 3 discusses the results of this review. Section 4 summarises the conclusions of this review.

DRN IN STLF

BASIC DRN STRUCTURE

The complex nonlinear interaction between the input data and the output is deciphered using the DRN (He et al., 2016). In general, as model depth increases, so does a neural network's capacity for learning. Ironically, though, this depth may work against the deep learning model. Either the inherent complexity of the data or the advanced nature of the model itself may be to blame for this performance drop. The architecture incorporates leftover blocks to overcome this difficulty. The learning process in these blocks involves mapping input to a residual function rather than immediately translating input to output. By improving the learning process through residual connections, guaranteeing improved gradient flow, and lowering the possibility of disappearing gradients, this method makes it easier to train deeper networks effectively. A residual network (ResNet) has two consecutive levels connected by a skip link, as shown in Figure 2.

In most cases, a skip connection functions as an identical mapping when the input and output dimensions match. Equation 1 represents the output of the relevant ResNet under these circumstances:

$$Y_{\text{output}} = X_{\text{Input}} + \mathcal{F}(X_{\text{Input}}, \Theta) \quad [1]$$

Nevertheless, the skip connection takes on the function of a linear projection when the input and output dimensions are different. In these cases, as described in Equation 2, the corresponding ResNet produces an output that integrates this linear projection (L_p).

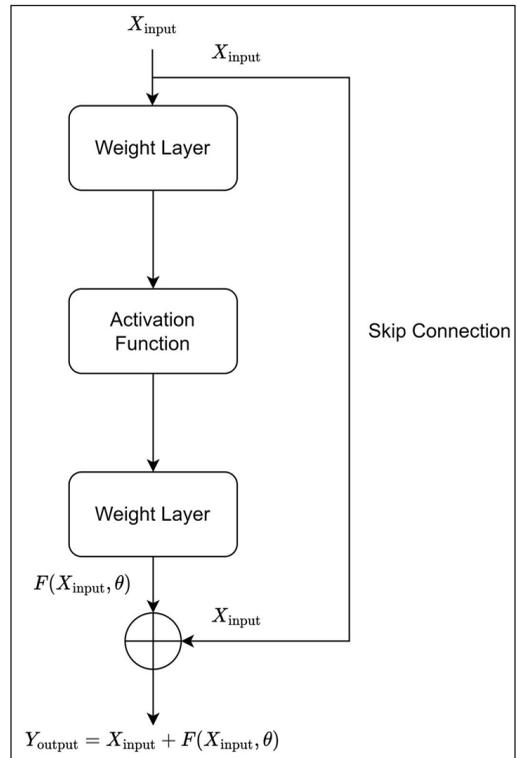


Figure 2. The basic component of the DRN (Chen et al., 2018)

$$y_{\text{output}} = L_p * x_{\text{Input}} + \mathcal{F}(x_{\text{Input}}, \theta) \quad [2]$$

A DRN may be easily constructed by stacking many residual blocks (ResBlocks). A DRN's structure is shown in Figure 3.

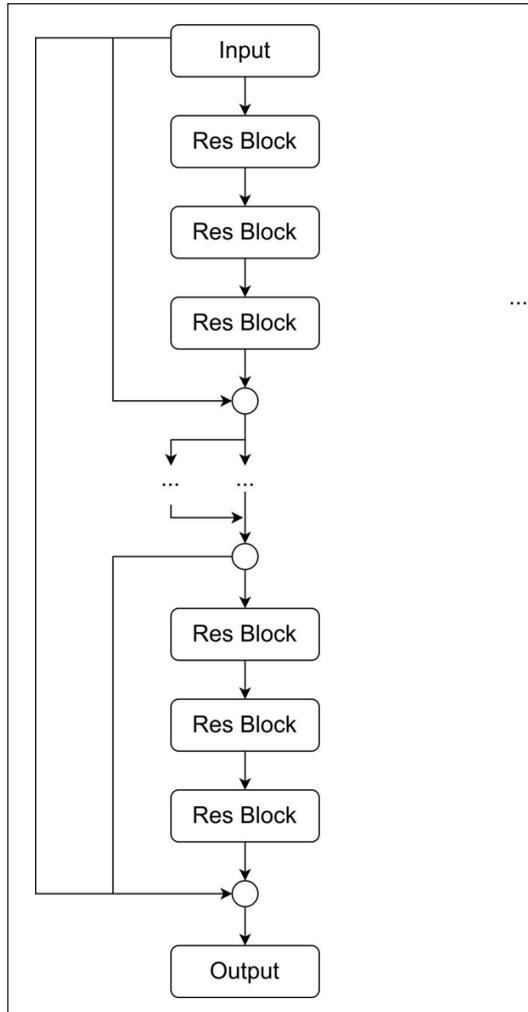


Figure 3. The structure of a DRN (Chen et al., 2018)

This skip link shows that a ResBlock's learning ability is at least comparable to that of a stack with a similar number of layers. The following is the forward-propagation Equation 3 where ResBlocks are stacked sequentially:

$$y(x) = x_0 + \sum_{j=1}^n F(x_j - 1, \theta_j - 1) \quad [3]$$

where x_0 is the input of the ResNet, x_n the output of the ResNet, and $\Theta_j = \{\theta_{j,z} | 1 \leq z \leq Z\}$ the set of weights associated with the j th ResBlock, Z being the number of layers within the block.

Equation 4 may then be used to compute the back propagation of the neural network's total loss to x_0 :

$$\frac{\partial \text{loss}}{\partial x_0} = \frac{\partial \text{loss}}{\partial x_n} \left(1 + \frac{\partial}{\partial x_0} \sum_{j=1}^n F(x_j - 1, \Theta_j - 1) \right) \quad [4]$$

The neural network's overall loss is represented by the provided equation. When "1" is present, it means that gradients from the network's output can be transmitted backward to its input directly. This straight backpropagation improves the network's learning efficiency by lowering the possibility of gradient disappearing, a typical problem when gradients must pass through several layers before reaching the input.

DRN STRUCTURE FOR STLF

The DRN for STLF builds on the basic DRN architecture described in (Chen et al., 2018) by integrating a core structure with an enhanced ResNet variant termed ResNetPlus. ResNetPlus preserves the block structure of the original ResNet while including improvements to optimise performance. It was developed specifically to improve the precision of 24-hour load estimates.

Initially a neural network with layers that are closely coupled, known as the 'basic structure' is used. The initial load projection for the next 24 hours is produced using this core architecture. Figure 4 shows a graphic representation of the model using the basic structure. Within this architecture, every fully connected (FC) layer corresponding to $[L_h^{\text{day}}, T_h^{\text{day}}]$, $[L_h^{\text{week}}, T_h^{\text{week}}]$, $[L_h^{\text{month}}, T_h^{\text{month}}]$ and L_h^{hour} comprises 10 hidden nodes. On the other hand, the fully-connected layers associated with $[S, W]$ are equipped with 5 hidden nodes. Additionally, both fully connected layer 1, FC2, and the FC layer preceding L_h contain 10 hidden nodes. It's worth noting that all layers, except for the output layer, employ an activation function.

The basic structure of the model is used for preliminary forecasting of the next 24 hours. In this framework, L_h^{month} denotes the load values for the corresponding hour from the days 1, 2, and 3 months before the predicted day. L_h^{week} signifies the load values for the same hour from 1 to 8 weeks prior, and L_h^{day} corresponds to the loads of the same hour for each day of the preceding week. L_h^{hour} represents the load values for the same hour from the previous 24 hours.

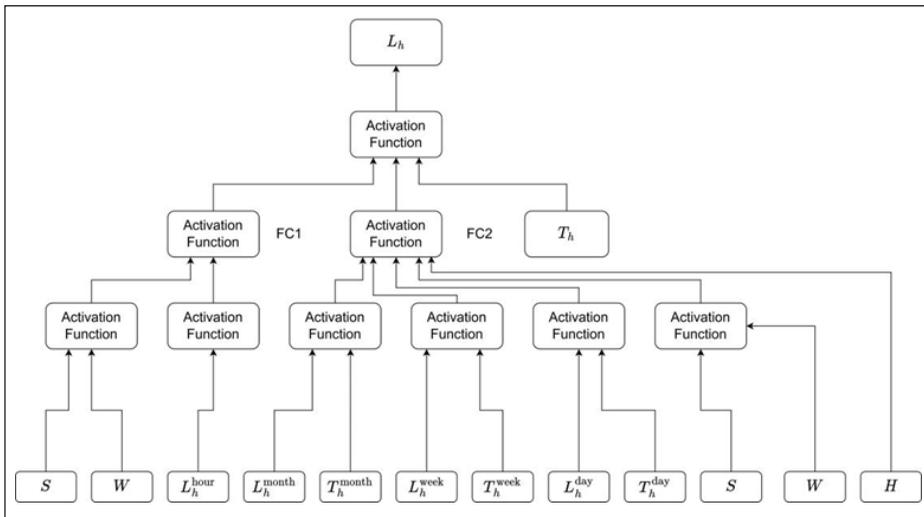


Figure 4. The basic structure for one-hour LF (Chen et al., 2018)

Additionally, T_h^{month} , T_h^{week} and T_h^{day} are the temperature readings concurrent with L_h^{month} , L_h^{week} , and L_h^{day} , respectively. T_h is the actual temperature forecasted for the next day. S , W , and H are one-hot encoded variables representing the season, weekday, and holiday status, respectively. The output from this basic structure, denoted as L_h , serves as the input for the second segment of the model, enhancing the forecasting precision.

With significant improvements over the original ResNet architecture, the ResNetPlus model is a sophisticated development in neural network architecture. This novel version incorporates ResBlocks with two hidden layers (20 nodes each) using the same activation methods. ResNetPlus generates a great deal of depth and complexity by successively constructing four of these blocks, each with its own unique connections, and repeating the process throughout ten levels. To reach the model's output, the design incorporates a unique shortcut link that runs straight from the last block's output to the network's entrance point. Such a configuration maximizes the effectiveness of a DRN while simultaneously making its development simpler. As seen in Figure 5, ResNetPlus enhances the original framework to fully use the residual design while maintaining the hyperparameters present in its ResNet predecessors inside these blocks.

To effectively train the models, the model's loss, denoted as $Loss$, is defined as the cumulative result of two distinct components Equation 5:

$$Loss = Loss_E + Loss_R \tag{5}$$

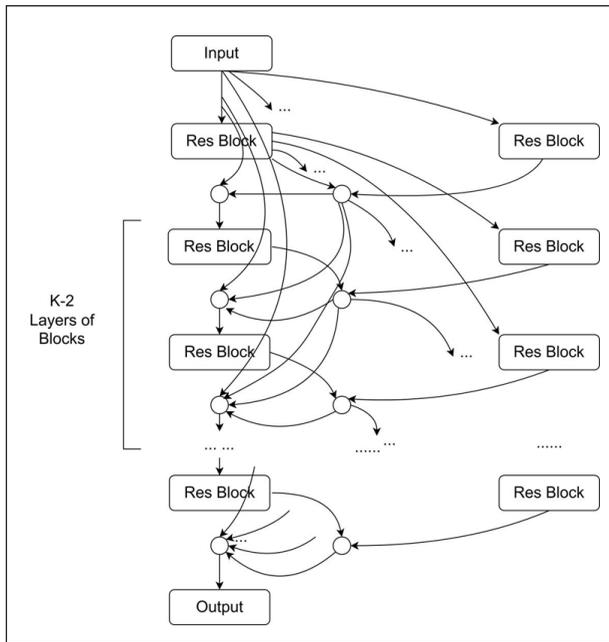


Figure 5. A depiction of the modified structure of the ResNetPlus (Chen et al., 2018)

Where $Loss_E$ quantifies the discrepancy in predictions, and $Loss_R$ serves as a penalty term for out-of-range values, designed to expedite the training phase. Particularly, $Loss_E$ is articulated as Equation 6:

$$Loss_E = \frac{1}{NumH} \sum_{j=1}^N \sum_{h=1}^H \frac{|\hat{y}_{(j,h)} - y_{(j,h)}|}{y_{(j,h)}} \tag{6}$$

where $\hat{y}_{(j,h)}$ represents the model's output and $y_{(j,h)}$ denotes the actual normalised load for the h th hour of the j th day. Here, Num symbolises the number of data samples, while H indicates the number of hourly loads within a day (notably, $H = 24$ in this scenario). This metric, commonly recognised as the Mean Absolute Percentage Error (MAPE), is employed both as a measure of error and as a criterion for assessing the forecast results of the models. The second term, $Loss_R$, is computed as Equation 7:

$$Loss_R = \frac{1}{2Num} \sum_{j=1}^{Num} \max(0, \max_h \hat{y}_{(j,h)} - \max_h y_{(j,h)}) + \max(0, \min_h y_{(j,h)} - \min_h \hat{y}_{(j,h)}) \tag{7}$$

$$\text{Loss}_R = \frac{1}{2\text{Num}} \sum_{j=1}^{\text{Num}} \max(0, \max_h \hat{y}_{(j,h)} - \max_h y_{(j,h)}) + \max(0, \min_h y_{(j,h)} - \min_h \hat{y}_{(j,h)}) \quad [7]$$

This term speeds up the first round of training by penalising the model if the expected daily load curves diverge from the actual load ranges. This phrase emphasises the cost of overestimating the load curves' peaks and underestimating their troughs when the model starts to produce forecasts with more precision.

OVERVIEW OF DRN APPLICATIONS IN STLF

Because of their capacity to simulate intricate and nonlinear load patterns, DRNs have shown advantages in STLF. These networks can overcome conventional deep learning obstacles like vanishing gradients thanks to the addition of residual connections, which also makes it possible to build deeper models that can learn richer feature representations. DRNs have therefore become a well-liked option for enhancing the precision, resilience, and dependability of LF models.

Establishing the viability of DRNs and contrasting them with current methods were the main goals of STLF's early use of them. The DRN model for STLF, an improved design that expanded on conventional ResNets by improving their capacity to manage nonlinear load patterns, was introduced by Chen et al. (Chen et al., 2018). By tackling overfitting and introducing advancements in feature extraction and generalisation, their study produced predictions that were more accurate. By introducing the ensemble ResNet (ERN), which combined several ResNets with ensemble learning approaches to improve generalisation, Xu et al. (2020) (Xu et al., 2020) advanced the discipline. This model demonstrated that even when exposed to noisy data, DRNs may retain a high level of predicted precision.

At about the same time, Tian et al. (Tian et al., 2020) investigated the combination of ResNetPlus and LSTMs, developing a hybrid strategy that capitalised on the LSTM's prowess in temporal pattern modelling and the DRN's ability to extract features. This invention demonstrated how integrating complementary architectures with DRNs might improve performance on intricate load datasets, a structure referred to as ResNetPlus-LSTM.

In 2021, DRN-based approaches became more diverse, building on the earlier work. When Kondaiah and Saravanan (Kondaiah & Saravanan, 2021) compared the DRN for STLF with the basic DRN structure, they found that the DRN for STLF design offered better prediction precision and generalisation. By integrating Google Net-inspired inception modules into a DRN framework, Ding et al. (Ding et al., 2021) went one step further. To increase computational efficiency and prediction precision, their Google Net-ResNetPlus model made use of multi-level residual connections and dimensionality reduction. To better

capture local patterns, Sheng et al. (Sheng et al., 2021) developed a convolutional ResNet (CRN) with unique convolutional blocks. The CRN showed notable gains in predicting precision and model stability when Leaky Rectified Linear Unit (Leaky ReLU) was used as the activation function, which made it a good fit for high-resolution load data.

Furthermore, Li et al. (Li et al., 2021) introduced a ResNet-LSTM-Attention model in 2021 that enhanced feature extraction and temporal pattern modelling by combining ResNets, LSTMs, and attention processes. Compared to conventional LSTM techniques, this integration improved precision and decreased MAPEs by enabling the model to preferentially focus on significant hidden states. The model's precision and stability across a range of datasets were further enhanced by the application of multi-stage training and Rectified Linear Unit (ReLU) activation.

More intricate hybrid and ensemble techniques emerged in the years that followed. Chen et al. (Chen et al., 2022) created an ensemble model based on ResNet that integrated a snapshot-based ensemble approach with multi-scale feature extraction. This method not only increased precision but also demonstrated the advantages of combining multiple perspectives on the data, a strategy referred to as the ResNet-Based Ensemble Model. The deep-ResNet model was first presented by Kondaiah and Saravanan (Kondaiah & Saravanan, 2022), who placed an emphasis on sophisticated feature selection methods to improve generalisation. The deep-ResNet model produced significant precision increases by employing Scaled Exponential Linear Unit (SELU) activation and methodically choosing input characteristics.

The Residual LSTM Plus model, which combined DRN blocks with LSTM layers, was introduced by Sheng et al. (Sheng et al., 2023) in 2023. This hybrid technique included the temporal modelling skills of the LSTM while maintaining the deep learning qualities of the DRN. This allowed it to achieve remarkable precision on datasets that needed to recognise both temporal and depth patterns.

A thorough summary of important research using DRN-based techniques for STLF is given in Table 1. The methods, temporal granularity, dataset properties, performance measurements, and activation functions used in these investigations differ. For example, early studies like Chen et al. (Chen et al., 2018) and Xu et al. (Xu et al., 2020) showed how to handle nonlinear patterns and increase prediction precision by concentrating on fundamental DRN structures such as ensemble ResNets and ResNetPlus. Hybrid approaches were introduced in later studies, including combining ResNetPlus with LSTMs (Tian et al., 2020), or augmenting DRN frameworks with inception modules (Ding et al., 2021) and convolutional ResBlocks (Sheng et al., 2021). Interestingly, the CRN created by Sheng et al. (Sheng et al., 2021) used Leaky ReLU as the activation function, demonstrating notable improvements in capturing local patterns. The evolution of these models shows a gradual increase in architectural complexity and an expansion of input features, allowing

Table 1
Comparison of DRN methods for STLF

References	Algorithm	Temporal granularity	Dataset variables	Data size	Metric	Activation function
Chen et al., 2018	DRN	Hourly	Load, time, temperature	North American Utility dataset, ISO-NE dataset	MAPE	SELU
Xu et al., 2020	ERN	Hourly	Load, time, temperature	North American Utility dataset, ISO-NE dataset	MAPE, RMSE, MAE, NMSE, R	SELU
Tian et al., 2020	ResNetPlus-LSTM	Hourly	Load, time, temperature	Local grid data (2006–2014)	MAPE, RMSE, MAE, MSE	SELU
Kondaiah & Saravanan, 2021	DRN	Hourly	Load, time, temperature	North American Utility dataset	MAPE	SELU
Ding et al., 2021	GoogleNet-ResNetPlus	Hourly	Load, time, temperature	North American Utility dataset, ISO-NE dataset	MAPE, RMSE, MAE, R ²	ReLU
Sheng et al., 2021	CRN	Hourly	Load, time, temperature	ISO-NE dataset, Malaysia dataset	MAPE	Leaky ReLU
Li et al., 2021	ResNet-LSTM-Attention	Hourly and half-hourly	Load, time, temperature, humidity	Hourly data from GEF2017, half-hourly data from LCL	MAPE, RMSE, MAE	ReLU
Chen et al., 2022	ResNet-Based Ensemble Model	Hourly	Load, time, temperature	North American Utility dataset, ISO-NE dataset	MAPE, RMSE, MAE	SELU
Kondaiah & Saravanan, 2022	deep-ResNet	Hourly	Load, time, temperature	ISO-NE dataset, IESO-Canada dataset	MAPE, RMSE, MAE	SELU
Sheng et al., 2023	Residual LSTM Plus	Hourly	Load, time, temperature	ISO-NE dataset, Malaysia dataset	MAPE, R ²	SELU

Table 2
Advantages and limitations of DRN methods for STLF

References	Algorithm	Advantages	Limitations
Chen et al., 2018	DRN	Mitigates vanishing gradients, enhances prediction precision	Limited long-term dependency modeling
Xu et al., 2020	ERN	Improved generalisation through ensemble learning	High computational complexity, reliance on snapshot models
Tian et al., 2020	ResNetPlus-LSTM	Combines short-term feature extraction with temporal modeling	Computationally intensive, limited variable diversity
Kondaiah & Saravanan, 2021	DRN	Improved precision and generalisation compared to standard DRNs	No new methods proposed, focuss limited to comparative analysis
Ding et al., 2021	Google Net-ResNetPlus	Reduces overfitting, enhances feature extraction	High computational cost, limited handling of long-term dependencies
Sheng et al., 2021	CRN	Excels in capturing local features, superior precision	Limited scalability for long-term forecasting, lacks diverse meteorological input integration
Li et al., 2021	ResNet-LSTM-Attention	Combines residual learning, LSTMs, and attention for precision	Does not explore long-term trends or multiple dataset scenarios
Chen et al., 2022	ResNet-Based Ensemble Model	High precision due to multi-scale feature extraction	Reliance on existing architectures, limited exploration of new variables or adaptive methods
Kondaiah & Saravanan, 2022	deep-ResNet	Strong generalisation and effective feature selection	Does not systematically address long-term dependencies or diverse input variables
Sheng et al., 2023	Residual LSTM Plus	Leverages LSTM for temporal patterns, DRN for depth	High computational complexity, limited exploration of variable diversity

DRNs to handle a variety of forecasting scenarios. Performance metrics (MAPE, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Normalised Mean Squared Error (NMSE), Correlation Coefficient (R), and Coefficient of Determination (R^2)), mainly MAPE, RMSE, and MAE, consistently highlight the precision improvements brought by DRNs, with SELU and ReLU being the most commonly used activation functions, and the occasional use of Leaky ReLU for capturing local features.

The primary benefits and drawbacks of DRN techniques in STLF are shown in Table 2. DRNs are strong because they can reduce overfitting, improve generalisation, and represent complex load patterns. Numerous studies demonstrate how they outperform conventional techniques in terms of precision, feature extraction efficiency, and stability. But there is a price for these advantages. High resource requirements, restricted scalability to long-term forecasting jobs, and computational complexity are common issues with DRNs. Furthermore, whereas DRNs are excellent at managing short-term dependencies, they become less successful when integrating several external factors or dealing with long-term patterns. These drawbacks highlight the necessity of continuous improvement and investigation of novel DRN designs and techniques.

DISCUSSION

This review highlights several areas where the potential of DRNs in STLF can be further refined. While DRNs have shown substantial success in capturing complex load patterns, key challenges remain.

First, the role of activation functions warrants more systematic investigation. Current studies rely heavily on standard choices such as ReLU and SELU yet fail to fully explore how alternative functions like Leaky ReLU, which has been used in certain CRN approaches might enhance convergence, improve generalisation, and reduce overfitting. Future work could provide a more comprehensive analysis of different activation functions, comparing their effects on model performance and identifying optimal configurations tailored to specific forecasting scenarios.

Second, while DRNs have made progress in addressing short-term dependencies, they still struggle with capturing long-term temporal patterns and global trends within load data. Current DRN architectures, often relying on fixed weights in FC layers, excel at identifying local, short-term features but fall short in dynamically learning long-range dependencies. Methods that integrate sequential modelling components, such as LSTM, have partially alleviated these limitations by extending the effective temporal window. However, these approaches often depend heavily on adjacent time steps and face difficulties in leveraging information across distant time points. The unidirectional nature of standard LSTM hinders the ability to fully utilise multi-cycle trends and

seasonality. As a result, the extracted long-term features remain incomplete, leading to underperformance when modelling data with pronounced periodic or trend-based characteristics. Future work should focus on refining DRN frameworks by integrating more advanced sequence modelling components to better capture complex long-term dependencies and global temporal patterns.

Lastly, the diversity of input variables remains limited. Most DRN studies primarily use load, time, and temperature data, leaving room for improvement by integrating additional variables, such as precipitation, humidity, or event-based factors. Advanced feature selection techniques, such as Principal Component Analysis (PCA), can help identify the most impactful variables. Expanding the scope of input features would allow DRNs to become more robust and adaptive, enabling them to respond more effectively to sudden changes in weather or load conditions.

CONCLUSION

DRNs have shown remarkable potential in STLF, demonstrating significant improvements in precision, feature extraction, and robustness. This review highlights both the key advancements and the persistent challenges associated with these models. DRNs have successfully captured local, short-term dependencies, but their ability to model long-term temporal patterns and global trends remains limited. Current approaches, including the integration of sequential modelling techniques such as LSTMs and attention mechanisms, have made progress in addressing these shortcomings but still fall short of fully leveraging distant temporal relationships and multi-cycle trends.

Future research should focus on further refining DRN architectures to strengthen their ability to model complex long-term dependencies. This includes exploring novel sequential modelling frameworks, enhancing attention-based mechanisms, and developing more sophisticated approaches to extract global patterns from load data. By addressing these challenges, DRNs can be transformed into even more powerful tools, delivering greater predictive precision and robustness in STLF tasks.

ACKNOWLEDGEMENT

Conceptualisation, Junchen Liu and Faisul Arif Ahmad; Methodology, Junchen Liu; Investigation, Junchen Liu and Faisul Arif Ahmad; Writing - original draft preparation, Junchen Liu; Writing - review and editing, Junchen Liu, Faisul Arif Ahmad, Khairulmizam Samsudin, Fazirulhisyam Hashim, and Mohd Zainal Abidin Ab Kadir; Supervision, Faisul Arif Ahmad, Khairulmizam Samsudin, Fazirulhisyam Hashim, and Mohd Zainal Abidin Ab Kadir. All authors have read and agreed to the published version of the manuscript.

ABBREVIATION

ANNs	: Artificial Neural Networks
BiGRU	: Bidirectional Gated Recurrent Unit
BiLSTM	: Bidirectional Long Short-Term Memory
R ²	: Coefficient of Determination
CNNs	: Convolutional Neural Networks
R	: Correlation Coefficient
DNNs	: Deep Neural Networks
DRN	: Deep Residual Network
ERN	: Ensemble ResNet
ELM	: Extreme Learning Machines
FC	: Fully Connected
GRU	: Gated Recurrent Units
Leaky ReLU	: Leaky Rectified Linear Unit
LF	: Load Forecasting
LSTM	: Long Short-Term Memory
LTLF	: Long-Term Load Forecasting
MAE	: Mean Absolute Error
MAPE	: Mean Absolute Percentage Error
MSE	: Mean Squared Error
MTLF	: Medium-Term Load Forecasting
NMSE	: Normalised Mean Squared Error
PCA	: Principal Component Analysis
RBF	: Radial Basis Function
ReLU	: Rectified Linear Unit
RNN	: Recurrent Neural Networks
ResBlock	: Residual Block
ResNet	: Residual Network
RMSE	: Root Mean Square Error
SELU	: Scaled Exponential Linear Unit
STLF	: Short-Term Load Forecasting
SVR	: Support Vector Regression
VSTLF	: Very Short-Term Load Forecasting

REFERENCES

- Akhtaruzzaman, M., Hasan, M. K., Kabir, S. R., Abdullah, S. N. H. S., Sadeq, M. J., & Hossain, E. (2020). HSIC bottleneck-based distributed deep learning model for load forecasting in smart grid with a comprehensive survey. *IEEE Access*, 8, 222977-223008. <https://doi.org/10.1109/ACCESS.2020.3040083>

- Alberg, D., & Last, M. (2018). Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms. *Vietnam Journal of Computer Science*, 5, 241-249. <https://doi.org/10.1007/s40595-018-0119-7>
- Ali, M., Adnan, M., Tariq, M., & Poor, H. V. (2020). Load forecasting through estimated parameterised-based fuzzy inference system in smart grids. *IEEE Transactions on Fuzzy Systems*, 29(1), 156-165. <https://doi.org/10.1109/TFUZZ.2020.2986982>
- Atef, S., Nakata, K., & Eltawil, A. B. (2022). A deep bidirectional long short-term memory neural network-based methodology to enhance short-term electricity load forecasting for residential applications. *Computers & Industrial Engineering*, 170, Article 108364. <https://doi.org/10.1016/j.cie.2022.108364>
- Baul, A., Sarker, G. C., Sikder, P., Mozumder, U., & Abdelgawad, A. (2024). Data-driven short-term load forecasting for multiple locations: An integrated approach. *Big Data and Cognitive Computing*, 8(2), Article 12. <https://doi.org/10.3390/bdcc8020012>
- Bento, P., Pombo, J., Mariano, S., & do Rosário Calado, M. (2018). Short-term load forecasting using optimised LSTM networks via improved bat algorithm. In *2018 International Conference on Intelligent Systems (IS)* (pp. 351-357). <https://doi.org/10.1109/IS.2018.8710498>
- Cai, C., Tao, Y., Zhu, T., & Deng, Z. (2021). Short-term load forecasting based on deep learning bidirectional LSTM neural network. *Applied Sciences*, 11(17), Article 8129. <https://doi.org/10.3390/app11178129>
- Cai, M., Pipattanasomporn, M., & Rahman, S. (2019). Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy*, 236, 1078-1088. <https://doi.org/10.1016/j.apenergy.2018.12.042>
- Cecati, C., Kolbusz, J., Różycki, P., Siano, P., & Wilamowski, B. M. (2015). A novel RBF training algorithm for short-term electric load forecasting and comparative studies. *IEEE Transactions on Industrial Electronics*, 62(10), 6519-6529. <https://doi.org/10.1109/TIE.2015.2424399>
- Cen, S., & Lim, C. G. (2024). Multi-task learning of the PatchTCN-TST model for short-term multi-load energy forecasting considering indoor environments in a smart building. *IEEE Access*. Advance online publication. <https://doi.org/10.1109/ACCESS.2024.3355448>
- Ceperic, E., Ceperic, V., & Baric, A. (2013). A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems*, 28(4), 4356-4364. <https://doi.org/10.1109/TPWRS.2013.2269803>
- Chen, K., Chen, K., Wang, Q., He, Z., Hu, J., & He, J. (2018). Short-term load forecasting with deep residual networks. *IEEE Transactions on Smart Grid*, 10(4), 3943-3952. <https://doi.org/10.1109/TSG.2018.2844307>
- Chen, W., Han, G., Zhu, H., Liao, L., & Zhao, W. (2022). Deep ResNet-based ensemble model for short-term load forecasting in protection system of smart grid. *Sustainability*, 14(24), Article 16894. <https://doi.org/10.3390/su142416894>
- Chen, Y., Luh, P. B., Guan, C., Zhao, Y., Michel, L. D., Coolbeth, M. A., & Rourke, S. J. (2009). Short-term load forecasting: Similar day-based wavelet neural networks. *IEEE Transactions on Power Systems*, 25(1), 322-330. <https://doi.org/10.1109/TPWRS.2009.2030426>
- Deng, Z., Wang, B., Xu, Y., Xu, T., Liu, C., & Zhu, Z. (2019). Multi-scale convolutional neural network with time cognition for multi-step short-term load forecasting. *IEEE Access*, 7, 88058-88071. <https://doi.org/10.1109/ACCESS.2019.2926137>

- Ding, A., Liu, T., & Zou, X. (2021). Integration of ensemble GoogLeNet and modified deep residual networks for short-term load forecasting. *Electronics*, *10*(20), Article 2455. <https://doi.org/10.3390/electronics10202455>
- Dong, X., Qian, L., & Huang, L. (2017). Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 119-125). <https://doi.org/10.1109/BIGCOMP.2017.7881726>
- Dudek, G. (2012). Variable selection in the kernel regression based short-term load forecasting model. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, & J. M. Zurada (Eds.), *Artificial intelligence and soft computing* (pp. 557-563). Springer. https://doi.org/10.1007/978-3-642-29350-4_66
- Eren, Y., & Küçükdemiral, İ. (2024). A comprehensive review on deep learning approaches for short-term load forecasting. *Renewable and Sustainable Energy Reviews*, *189*, Article 114031. <https://doi.org/10.1016/j.rser.2023.114031>
- Ertugrul, Ö. F. (2016). Forecasting electricity load by a novel recurrent extreme learning machines approach. *International Journal of Electrical Power & Energy Systems*, *78*, 429-435. <https://doi.org/10.1016/j.ijepes.2015.12.006>
- Fan, S., & Hyndman, R. J. (2011). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, *27*(1), 134-141. <https://doi.org/10.1109/TPWRS.2011.2162082>
- Fang, B., Xu, L., Luo, Y., Luo, Z., & Li, W. (2024). A method for short-term electric load forecasting based on the FMLP-iTransformer model. *Energy Reports*, *12*, 3405-3411. <https://doi.org/10.1016/j.egy.2024.09.023>
- Gan, D., Wang, Y., Zhang, N., & Zhu, W. (2017). Enhancing short-term probabilistic residential load forecasting with quantile long short-term memory. *The Journal of Engineering*, *2017*(14), 2622-2627. <https://doi.org/10.1049/joe.2017.0833>
- Giacomazzi, E., Haag, F., & Hopf, K. (2023). Short-term electricity load forecasting using the temporal fusion transformer: Effect of grid hierarchies and data sources. In *Proceedings of the 14th ACM International Conference on Future Energy Systems* (pp. 353-360). ACM. <https://doi.org/10.1145/3575813.3597345>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Han, L., Peng, Y., Li, Y., Yong, B., Zhou, Q., & Shu, L. (2018). Enhanced deep networks for short-term and medium-term load forecasting. *IEEE Access*, *7*, 4045-4055. <https://doi.org/10.1109/ACCESS.2018.2888978>
- Haque, A., & Rahman, S. (2022). Short-term electrical load forecasting through heuristic configuration of regularized deep neural network. *Applied Soft Computing*, *122*, Article 108877. <https://doi.org/10.1016/j.asoc.2022.108877>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). <https://doi.org/10.1109/CVPR.2016.90>
- He, W. (2017). Load forecasting via deep neural networks. *Procedia Computer Science*, *122*, 308-314. <https://doi.org/10.1016/j.procs.2017.11.374>
- Hernández, L., Baladrón, C., Aguiar, J. M., Carro, B., Sánchez-Esguevillas, A., & Lloret, J. (2014). Artificial neural networks for short-term load forecasting in microgrids environment. *Energy*, *75*, 252-264. <https://doi.org/10.1016/j.energy.2014.07.065>

- Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44-55. <https://doi.org/10.1109/59.910780>
- Huang, S., Zhang, J., He, Y., Fu, X., Fan, L., Yao, G., & Wen, Y. (2022). Short-term load forecasting based on the CEEMDAN-sample entropy-BPNN-transformer. *Energies*, *15*(10), Article 3659. <https://doi.org/10.3390/en15103659>
- Huy, P. C., Minh, N. Q., Tien, N. D., & Anh, T. T. Q. (2022). Short-term electricity load forecasting based on temporal fusion transformer model. *IEEE Access*, *10*, 106296-106304. <https://doi.org/10.1109/ACCESS.2022.3211941>
- Jiang, B., Liu, Y., Geng, H., Zeng, H., & Ding, J. (2022). A transformer-based method with wide attention range for enhanced short-term load forecasting. In *2022 4th International Conference on Smart Power & Internet Energy Systems (SPIES)* (pp. 1684-1690). <https://doi.org/10.1109/SPIES55999.2022.10082249>
- Jurado, M., Samper, M., & Rosés, R. (2023). An improved encoder-decoder-based CNN model for probabilistic short-term load and PV forecasting. *Electric Power Systems Research*, *217*, Article 109153. <https://doi.org/10.1016/j.epsr.2023.109153>
- Kondaiah, V. Y., & Saravanan, B. (2021). Short-term load forecasting with deep learning. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-5). <https://doi.org/10.1109/i-PACT52855.2021.9696634>
- Kondaiah, V. Y., & Saravanan, B. (2022). A modified deep residual network for short-term load forecasting. *Frontiers in Energy Research*, *10*, Article 1038819. <https://doi.org/10.3389/fenrg.2022.1038819>
- Kondaiah, V. Y., Saravanan, B., Sanjeevikumar, P., & Khan, B. (2022). A review on short-term load forecasting models for micro-grid application. *The Journal of Engineering*, *2022*(7), 665-689. <https://doi.org/10.1049/tje2.12151>
- Kong, W., Dong, Z. Y., Hill, D. J., Luo, F., & Xu, Y. (2017). Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems*, *33*(1), 1087-1088. <https://doi.org/10.1109/TPWRS.2017.2688178>
- Kong, Z., Zhang, C., Lv, H., Xiong, F., & Fu, Z. (2020). Multimodal feature extraction and fusion deep neural networks for short-term load forecasting. *IEEE Access*, *8*, 185373-185383. <https://doi.org/10.1109/ACCESS.2020.3029828>
- Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, *35*, 257-270. <https://doi.org/10.1016/j.scs.2017.08.009>
- Kwon, B. S., Park, R. J., & Song, K. B. (2020). Short-term load forecasting based on deep neural networks using LSTM layer. *Journal of Electrical Engineering & Technology*, *15*, 1501-1509. <https://doi.org/10.1007/s42835-020-00424-7>
- Li, H., Zhang, P., & Li, C. (2021). Short-term load forecasting for distribution substations based on residual neural networks and long short-term memory neural networks with attention mechanism. *Journal of Physics: Conference Series*, *2030*(1), Article 012087. <https://doi.org/10.1088/1742-6596/2030/1/012087>
- Li, L., Ota, K., & Dong, M. (2017). Everything is image: CNN-based short-term electrical load forecasting for smart grid. In *Proceedings of the 2017 14th International Symposium on Pervasive Systems, Algorithms*

- and Networks (ISPAN), the 2017 11th International Conference on Frontier of Computer Science and Technology (FCST), and the 2017 Third International Symposium of Creative Computing (ISCC) (pp. 344-351). <https://doi.org/10.1109/ISPAN-FCST-ISCC.2017.78>
- Li, S., Zhang, W., & Wang, P. (2023). TS2ARCformer: A multi-dimensional time series forecasting framework for short-term load prediction. *Energies*, *16*(15), Article 5825. <https://doi.org/10.3390/en16155825>
- Liao, H., & Radhakrishnan, K. K. (2022). Short-term load forecasting with temporal fusion transformers for power distribution networks. In *2022 IEEE Sustainable Power and Energy Conference (iSPEC)* (pp. 1-5). <https://doi.org/10.1109/iSPEC54162.2022.10033079>
- Lin, Z., Xie, L., & Zhang, S. (2024). A compound framework for short-term gas load forecasting combining time-enhanced perception transformer and two-stage feature extraction. *Energy*, *298*, Article 131365. <https://doi.org/10.1016/j.energy.2024.131365>
- Liu, J., Ahmad, F. A., Samsudin, K., Hashim, F., & Ab Kadir, M. Z. A. (2025). Performance evaluation of activation functions in deep residual networks for short-term load forecasting. *IEEE Access*, *13*, 78618-78633. <https://doi.org/10.1109/ACCESS.2025.3565798>
- Liu, M., Li, Y., Hu, J., Wu, X., Deng, S., & Li, H. (2023). A new hybrid model based on SCINet and LSTM for short-term power load forecasting. *Energies*, *17*(1), Article 95. <https://doi.org/10.3390/en17010095>
- Liu, Q., Cao, J., Zhang, J., Zhong, Y., Ba, T., & Zhang, Y. (2023). Short-term power load forecasting in FGSM-Bi-LSTM networks based on empirical wavelet transform. <https://doi.org/10.1109/ACCESS.2023.3316516>
- Lu, W., & Chen, X. (2024). Short-term load forecasting for power systems with high-penetration renewables based on multivariate data slicing transformer neural network. *Frontiers in Energy Research*, *12*, Article 1355222. <https://doi.org/10.3389/fenrg.2024.1355222>
- Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. In *IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society* (pp. 7046-7051). <https://doi.org/10.1109/IECON.2016.7793413>
- Muzaffar, S., & Afshari, A. (2019). Short-term load forecasts using LSTM networks. *Energy Procedia*, *158*, 2922-2927. <https://doi.org/10.1016/j.egypro.2019.01.952>
- Narayan, A., & Hipel, K. W. (2017). Long short-term memory networks for short-term electric load forecasting. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2573-2578). <https://doi.org/10.1109/SMC.2017.8123012>
- Ran, P., Dong, K., Liu, X., & Wang, J. (2023). Short-term load forecasting based on CEEMDAN and transformer. *Electric Power Systems Research*, *214*, Article 108885. <https://doi.org/10.1016/j.epsr.2022.108885>
- Rejc, M., & Pantos, M. (2011). Short-term transmission-loss forecast for the Slovenian transmission power system based on a fuzzy-logic decision approach. *IEEE Transactions on Power Systems*, *26*(3), 1511-1521. <https://doi.org/10.1109/TPWRS.2010.2096829>
- Santos, M. L., García, S. D., García-Santiago, X., Ogando-Martínez, A., Camarero, F. E., Gil, G. B., & Ortega, P. C. (2023). Deep learning and transfer learning techniques applied to short-term load forecasting of data-poor buildings in local energy communities. *Energy and Buildings*, *292*, Article 113164. <https://doi.org/10.1016/j.enbuild.2023.113164>

- Sheng, Z., An, Z., Wang, H., Chen, G., & Tian, K. (2023). Residual LSTM-based short-term load forecasting. *Applied Soft Computing*, 144, Article 110461. <https://doi.org/10.1016/j.asoc.2023.110461>
- Sheng, Z., Wang, H., Chen, G., Zhou, B., & Sun, J. (2021). Convolutional residual network for short-term load forecasting. *Applied Intelligence*, 51(4), 2485-2499. <https://doi.org/10.1007/s10489-020-01932-9>
- Shi, Y., Wang, S., Zhang, L., Yang, F., & Luo, X. (2023). A novel short-term power-load forecasting method based on high-dimensional meteorological data dimensionality reduction and hybrid deep neural network. *Journal of Energy Engineering*, 149(6), Article 04023049. <https://doi.org/10.1061/JLEED9.EYENG-5009>
- Song, K. B., Baek, Y. S., Hong, D. H., & Jang, G. (2005). Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Transactions on Power Systems*, 20(1), 96-101. <https://doi.org/10.1109/TPWRS.2004.835632>
- Tang, X., Dai, Y., Liu, Q., Dang, X., & Xu, J. (2019). Application of bidirectional recurrent neural network combined with deep belief network in short-term load forecasting. *IEEE Access*, 7, 160660-160670. <https://doi.org/10.1109/ACCESS.2019.2950957>
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8), 799-805. <https://doi.org/10.1057/palgrave.jors.2601589>
- Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1), 139-152. <https://doi.org/10.1016/j.ejor.2009.10.003>
- Tian, C., Ma, J., Zhang, C., & Zhan, P. (2018). A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies*, 11(12), Article 3493. <https://doi.org/10.3390/en11123493>
- Tian, Y., Yu, S., Wen, M., Zhang, K., & Chen, Y. (2020). Short-term load forecasting scheme based on improved deep residual network and LSTM. In *CIREC 2020 Berlin Workshop (CIREC 2020)* (pp. 117-120). <https://doi.org/10.1049/oap-cired.2021.0257>
- ul Hassnain, S. R., & Khan, A. (2007). Short-term load forecasting using particle swarm optimisation-based ANN approach. In *2007 International Joint Conference on Neural Networks* (pp. 1476-1481). <https://doi.org/10.1109/IJCNN.2007.4371176>
- Upadhyay, A., Garg, D., & Singh, M. (2023). Short-term load forecasting for smart grids using Apache Spark and a modified transformer model. *Computing and Informatics*, 42(1), 75-97. https://doi.org/10.31577/cai_2023_1_75
- Valencia, V. A. N., & Sanchez-Galan, J. E. (2022). Use of attention-based neural networks for short-term load forecasting in the Republic of Panama. In *2022 IEEE 40th Central America and Panama Convention (CONCAPAN)* (pp. 1-6). <https://doi.org/10.1109/CONCAPAN48024.2022.9997752>
- Wang, S., Wang, X., Wang, S., & Wang, D. (2019). Bidirectional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, 109, 470-479. <https://doi.org/10.1016/j.ijepes.2019.02.022>
- Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., & Kang, C. (2019). Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy*, 235, 10-20. <https://doi.org/10.1016/j.apenergy.2018.10.078>

- Wu, W., Liao, W., Miao, J., & Du, G. (2019). Using gated recurrent unit network to forecast short-term load considering impact of electricity price. *Energy Procedia*, 158, 3369-3374. <https://doi.org/10.1016/j.egypro.2019.01.950>
- Xu, C., & Chen, G. (2024). Interpretable transformer-based model for probabilistic short-term forecasting of residential net load. *International Journal of Electrical Power & Energy Systems*, 155, Article 109515. <https://doi.org/10.1016/j.ijepes.2023.109515>
- Xu, Q., Yang, X., & Huang, X. (2020). Ensemble residual networks for short-term load forecasting. *IEEE Access*, 8, 64750-64759. <https://doi.org/10.1109/ACCESS.2020.2984722>
- Zhang, G., Wei, C., Jing, C., & Wang, Y. (2022). Short-term electrical load forecasting based on time-augmented transformer. *International Journal of Computational Intelligence Systems*, 15(1), Article 67. <https://doi.org/10.1007/s44196-022-00128-y>
- Zhang, Q., Chen, J., Xiao, G., He, S., & Deng, K. (2023). Transform Graph: A novel short-term electricity net load forecasting model. *Energy Reports*, 9, 2705-2717. <https://doi.org/10.1016/j.egypr.2023.01.050>
- Zhao, Y., Luh, P. B., Bomgardner, C., & Beerel, G. H. (2009). Short-term load forecasting: Multi-level wavelet neural networks with holiday corrections. In *2009 IEEE Power & Energy Society General Meeting* (pp. 1-7). <https://doi.org/10.1109/PES.2009.5275304>
- Zhao, Z., Xia, C., Chi, L., Chang, X., Li, W., Yang, T., & Zomaya, A. Y. (2021). Short-term load forecasting based on the transformer model. *Information*, 12(12), Article 516. <https://doi.org/10.3390/info12120516>