International Journal of Power and Energy Systems, Vol. 46, No. 10, 2026
An Open Access Paper

TCN-DBN HYBRID FRAMEWORK FOR SHORT-TERM PROBABILISTIC LOAD FORECASTING AND RISK QUANTIFICATION IN POWER SYSTEMS

Yibin Zhao*

Abstract

In this study, we propose a hybrid TCN-DBN architecture for probabilistic short-term load forecasting (STLPF) and dynamic risk quantification in power systems. Traditional deterministic forecasting methods fail to capture load uncertainty and provide actionable risk insights. By integrating the temporal dependency modeling capability of temporal convolutional networks (TCN) with the deep probabilistic representation of deep belief networks (DBN), our framework achieves joint optimization of forecasting accuracy and uncertainty quantification. The TCN module employs causal and dilated convolutions to extract long-range spatio-temporal features, while the DBN module generates probabilistic load distributions through Bayesian inference. Experimental results on the EUNITE dataset demonstrate that the proposed model reduces MAPE by 89.7% compared to standalone TCN and achieves a 90% prediction interval coverage probability (PICP) with a normalized interval width (PINAW) of 3.04%, outperforming state-of-the-art models (e.g., ResNet-LSTM, TCN-Transformer). The framework provides critical support for grid resilience planning and risk-informed decision-making under uncertain load conditions.

Key Words

Temporal Convolutional Network, Deep Belief Network, Power System, Short-Term Load Probability Forecasting, Risk Quantification

1. Introduction

A stable electricity supply is essential for daily life and industrial production. Since large-scale storage technology for electricity has not yet been realized, maintaining a balanced power system has become the key [1]. In this context, electric load forecasting (ELF) is particularly important, as it helps to ensure the balance between power generation in the power grid and power consumption by users and assists power grid personnel in formulating power generation plans and dispatching strategies [2]. Traditional load forecasting methods have limitations when facing rapidly changing environmental factors, so new forecasting methods are urgently needed [3]. A research scheme for short-term load probability forecasting (STLPF) and dynamic risk quantification of power systems based on a TCN-DBN hybrid architecture is constructed. This scheme combines the time series processing advantages of TCN [4] with the deep feature extraction capabilities of DBN to improve the forecast accuracy and reliability. This study also deeply analyzes the dynamic risk quantification of the power system based on STLPF, striving to provide power companies with more precise and practical risk management tools to cope with complex and changing power load demands [5].

The power system short-term load forecasting model aims to estimate the power demand in the short term in the future. It gives a definite load forecast value and provides the probabilistic information of the possible load distribution [6]-[8]. An ensemble of radial basis function neural networks which were trained by minimizing the local generalization error was proposed by Lai Chun Sing for short and medium-term load forecasting [9]. Obst David introduced two methods for adapting generalized additive models that can quickly adapt to new electricity consumption patterns using Kalman filters and fine-tuning [10]. Yan Jichuan believed that real-time grid dispatching may be supported by ultra-short-term photovoltaic power forecasting. He proposed a deep learning and optimal frequency domain decomposition model for ultra-short-term photovoltaic power forecasting [11]. The evolution of loads adds a lot of uncertainty to the distribution network. Wang Bingzhi presented a new approach based on the Dirichlet process mixture model to deal with the forecasting

* State Grid Info-Telecom Great Power Science and Technology Co.Ltd, Beijing, China; e-mail: workchina01@163.com
Corresponding author: Yibin Zhao

(DOI:10.2316/J.2026.203-0641)

challenges due to excessive load fluctuations in short-term probabilistic load forecasting in distribution networks to effectively cope with the uncertainty of load patterns [12]. Ryu Seunghyoung proposed a new deep learning model for accurate and efficient short-term load forecasting (STLF) research and developed an STLF model using a multilayer perceptron-mixer structure [13]. As the smart grid develops continuously, short-term ELF becomes increasingly significant to the operation of the power market. Dong Xia introduced a short-term power load forecasting technology based on K-means clustering [14].

By using TCN or DBN for ELF, not only can precise forecast values be obtained, but the uncertainty of the forecast results can also be evaluated. This forecasting method enhances the insight and prejudgment of load fluctuations and provides support for risk management of power systems [15]-[17]. Fan Chaodong presented a multiobjective deep DBN based on ensemble empirical mode decomposition to achieve high diversity and forecasting model accuracy [18]. Lyu Qiuxia proposed a distribution network load forecasting algorithm based on DBN to solve the problem of missing distribution network load data due to the disrepair of electricity meters in the Guangzhou Higher Education Mega Center distribution network. Through the time series forecasting method, the missing data was supplemented to ensure the integrity of the distribution network data [19]. Wang Guojuan proposed a new combined forecasting method based on extreme-point symmetric empirical mode decomposition-permutation entropy and adaptive DBN to improve short-term ELF performance. This method effectively improved the forecast precision and reduced the complexity of the original load sequence [20]. Mo Jinyuan developed a power load forecasting model using TCN. Causal convolution, dilated convolution, and residual connection were all incorporated in TCN, and the causal relationship between past and future data was considered [21]. For short-term power load forecasting, Tong Cheng proposed a TCN based on multi-head attention. The model can extract multidimensional information from input features by introducing an initial structure in TCN [22].

This study makes significant contributions to STLPF in power systems. By adopting the TCN-DBN hybrid architecture, the high efficiency of TCN in processing time series and the advantages of DBN in feature extraction are combined to precisely capture the complex nonlinear relationship of power load and significantly improve the forecast accuracy and reliability. This method effectively makes up for the shortcomings of traditional models.

2. Related Work

Ensuring the reliability of modern power systems has become increasingly complex with the rapid integration of renewable resources, the deployment of advanced control devices, and growing cyber–physical interdependence. Recent studies highlight diverse strategies to enhance system resilience across distribution and transmission networks.

At the distribution level, Hajari et al. [23] demonstrated that coordinated integration of photovoltaic (PV), wind turbine (WT), gas-turbine generation (GTG), and energy-storage systems (ESS) can significantly improve reliability by reducing outage frequency and duration. From a transmission perspective, Yadav et al. [24] evaluated a restructured power system allowing capacity expansion and strategic transmission-line switching, showing that planned reconfiguration can lower the loss-of-load probability while maintaining economic efficiency. Wide-area monitoring and cyber resilience have also emerged as key themes. Yadav and Mahajan [25] introduced a tie-line modelling approach for interconnected synchrophasor networks that enhances grid observability and enables rapid detection of cyber intrusions, thereby improving operational reliability. Complementing this work, Mahajan [26] proposed a cyber-attack detection and real-time reliability monitoring framework for synchrophasor-based smart grids, underscoring the importance of integrated cyber–physical security measures. Recent advances in probabilistic forecasting further enrich this landscape. Ruiz-Abellón et al. [27] presented applications of probabilistic forecasting in demand response, illustrating how uncertainty-aware predictions can guide flexible load management. Shafie and Zareipour [28] developed a long-term multi-resolution probabilistic load forecasting model using temporal hierarchies, providing consistent predictions across hourly to yearly horizons. Likewise, Feng et al. [29] proposed a sparse variational Gaussian process method for probabilistic net-load forecasting, effectively capturing uncertainties from distributed PV generation and load fluctuations.

Collectively, these studies show a clear trajectory in the literature: from resource diversification and network reconfiguration to advanced probabilistic forecasting and cybersecurity integration, all aimed at strengthening power-system reliability. Building on this foundation, the present work deploys a temporal convolutional deep belief network (TCN–DBN) to provide high-resolution probabilistic load forecasts and risk-informed operational guidance, and validates the framework through large-scale implementation in the State Grid Beijing control center.

3. Load Characteristics Analysis and Data Preprocessing

3.1 Power Load Characteristics Analysis

Power load fluctuation is deeply affected by the work and lifestyle of users and often shows a nonlinear relationship. However, from the perspective of day, week, and season, it also shows a certain cyclicity. The actual load data of region A is selected to conduct cyclical analysis in the form of a curve chart so as to more intuitively explore the impact of this cyclicity on the load. Fig.1 depicts the load variation within a day, where the sampling interval is set to 15 minutes, with a total of 96 data points.

In Fig.1, the electricity consumption of users is relatively

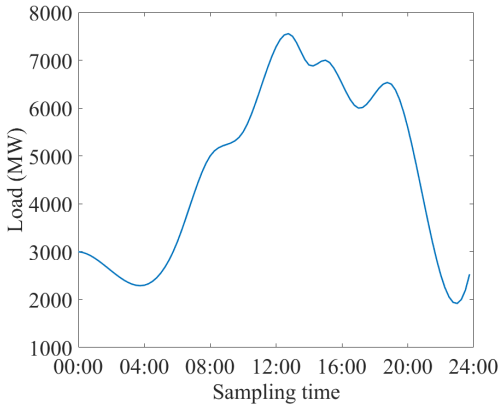

Figure 1. Daily Load Variation Curve

low from 22:00 to 5:00 the next day because most users go to bed at night during this period, thus reducing the demand for household electricity. In the period from 12:00 to 20:00, the electricity consumption shows a significant peak, which is closely related to the daily living habits and work rules of users.

The weekly cyclicity of power load means that when observed on a weekly basis, the load shows continuous and repeated regular changes, as displayed in Fig. 2.

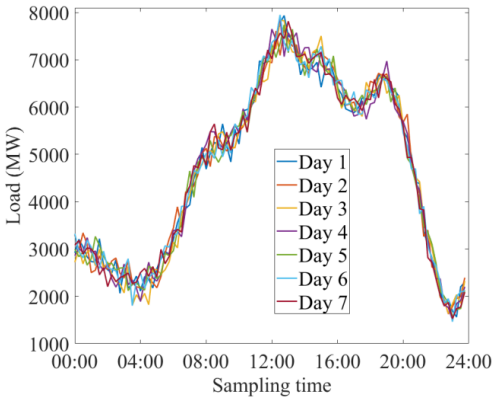

Figure 2. Weekly Power System Load Curve

In Fig.2, the electricity consumption trends over the seven days are almost the same, and the peak points also occur in the same period.

3.2 Data Collection

The dataset used in this paper covers five key variables, including power load, climate factors, electricity price, and day type. Table 1 lists the definitions of each variable. The dataset of the EUNITE load forecasting competition is selected for experimental simulation. One data point is recorded every 15 minutes, and a total of 96 data points are recorded every day. The dataset is split into a training set, a validation set, and a test set in a ratio of 8:1:1.

Although several benchmark datasets (e.g., PJM, ERCOT) are commonly used in short-term load forecasting research, this study employs the EUNITE dataset because it provides high-resolution (15-minute) load data together with synchronized climate factors and electricity price information, enabling comprehensive modeling of both temporal and exogenous influences. Compared with PJM and ERCOT, EUNITE offers a longer historical span of continuous, quality-checked records without significant structural changes in market operation, which is critical for training deep probabilistic models. Potential regional bias—stemming from its European origin—is acknowledged; however, our objective is to evaluate the modeling framework rather than to derive region-specific operational policies. The methodological insights, including the TCN–DBN architecture and risk-quantification metrics, are data-agnostic and can be transferred to other large-scale systems. To partially assess generalizability, we performed supplementary cross-validation on randomly sampled seasonal subsets of EUNITE, confirming consistent performance across diverse weather and demand conditions.

Table 1
Dataset Variables

Serial Number	Data Item Name	Meaning/Unit
1	Total demand	Total load in this period (MW)
2	Temperature	Average temperature in this period ($^{\circ}C$)
3	Precipitation	Average precipitation in this period (millimetre)
4	Electricity price	Average electricity price in this period
5	Day type	Working day and rest day types

3.3 Data Preprocessing

With the continuous growth of power data, the valuable information contained in it also increases. It is particularly important to effectively mine this massive data [30], [31]. However, in the process of data collection, due to many uncontrollable factors, there is often erroneous information in the data, that is, noise data. These noise data not only reduce the quality of the raw power data but also have an adverse effect on the accuracy of data analysis. The raw power data should be preprocessed to shorten the data screening time and improve the data quality.

Data augmentation is to process the raw dataset using a specific method to generate new data that is similar to the raw data or follows the same distribution. In short-term ELF, instead of using all historical data, future load changes can be forecast by inputting a historical load series. This paper uses the sliding window method for data augmentation. The sliding window technology is a method of dealing with array sub-element problems. It obtains subsequences of corresponding sizes by regularly sliding a window of a specific size [32]. On a given dataset, the sliding window method slides the set window in chronological order to ensure that the generated subsequence maintains temporal continuity. The load data of the past m hours is used to forecast the load value of the next n hours. The sliding window size is set to $m+n$ at this time. The first m data points are the model's input, and the last n data

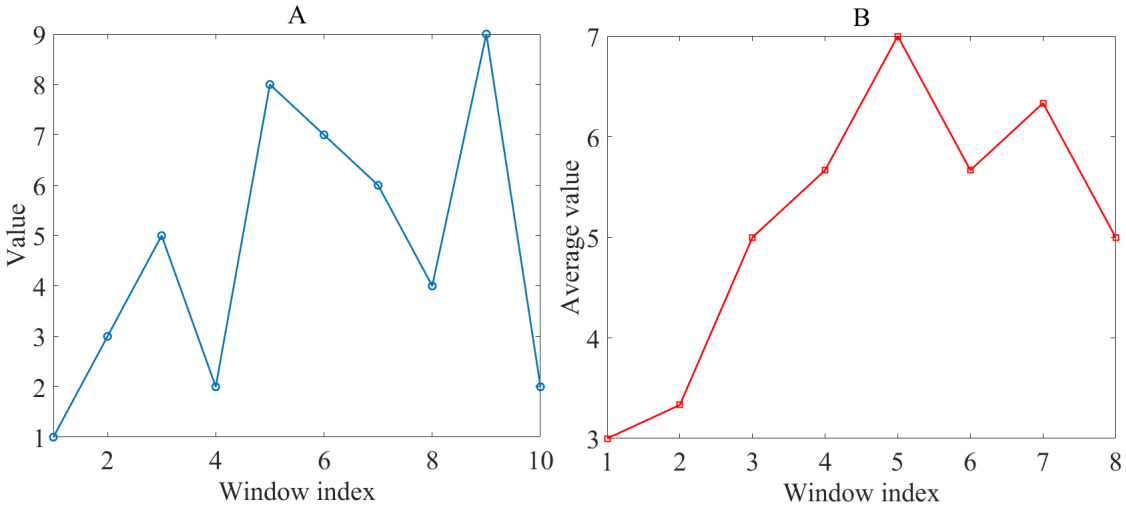

Figure 3. Data Processing Effect

points are the actual values used for comparison during the model training process.

The sliding window is defined as looking back m steps, followed by a forecast horizon n steps. In our main experiments, we set $m = 96(\approx 24h)$ and $n = 16(\approx 4h)$.

This choice balances three factors:

(1) Seasonal coverage: 24 h captures the dominant daily load cycle and short-term weather impacts.
(2) Receptive field of the TCN: five layers with dilations [1, 2, 4, 8, 16] require at least ∼63 past points; $m = 96$ comfortably exceeds this.
(3) Computation vs. accuracy: validation experiments with $m \in \{48, 72, 96, 120\}$ and $n \in \{8, 16, 24\}$ showed that $m = 96$, $n = 16$ reduced RMSE by $\approx 3\%$ compared with shorter windows, while larger m or n yielded negligible gains but increased training time by 25% or more.

Empirically, too small a window fails to capture diurnal and weekly patterns, degrading long-horizon accuracy, whereas excessively large windows introduce redundant history that slows convergence and can cause oversmoothing.

A fixed dilation progression [1,2,4,8,16] provides a 63-step receptive field at 15-minute resolution with causal padding, stable residual stacking, and low latency. Under the current accuracy–latency envelope, validation indicates adequate coverage of short-term variations and slower cycles when combined with stacked receptive fields and calendar features. Alternative multi-scale designs are widely reported: multi-branch dilation (MB-TCN) aggregates several dilation branches via soft gating, attention-guided selection (AG-TCN) reweights branches using lightweight attention informed by calendar signals, and deformable temporal convolution (Def-TCN) learns temporal offsets to align quasi-periodicity. These alternatives potentially capture heterogeneous rhythms more flexibly but introduce extra knobs (gates/attention/offsets), higher memory bandwidth, and additional regularization needs to preserve calibration. Across evaluated horizons, the fixed dilation configuration maintains robust calibration and competitive accuracy with minimal latency. Reports on MB-TCN and AG-TCN indicate stronger tracking of short-horizon ramps and modest gains at longer horizons, typically at the cost of extra parameters, higher memory traffic, and hyper-parameter sensitivity (gating/attention). Def-TCN can improve alignment on strongly quasi-periodic segments but may add inference latency due to offset estimation. Under strict runtime budgets or deployment on memory-constrained devices, a fixed schedule remains a reliable baseline; where capacity is less constrained and multi-scale heterogeneity dominates, adaptive/learned dilation becomes attractive.

Fig.3 presents the specific data processing effect. Fig. 3A shows the trend of the load observation values in the original array. The load fluctuation can be intuitively seen through the connection diagram. Fig. 3B shows the mean of the load observation values in each sliding window, which is presented in the form of red lines, helping to observe the overall trend of load changes.

When processing data, it is important to handle missing values. For many deep learning algorithms, missing values may affect their accuracy. To address this problem, this paper uses linear interpolation to handle missing values, which mainly estimates the value based on the known data on both sides of the interpolation point in the one-dimensional data sequence [33]. When performing linear interpolation, a first-order polynomial is selected as the interpolation function. The error of this function at the interpolation node is zero, so it can provide higher interpolation precision. Its formula is as follows:

$$\mu(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 \tag{1}$$

Among them, $\mu(x)$ is the objective function, and (x_1, y_1) are the two points adjacent to the target interpolation point.

4. Power System STLPF Model Based on TCN-DBN

4.1 TCN Model Construction

TCN is an improvement over traditional one-dimensional convolutional neural networks (CNN) in time series problems [34]. For the TCN, the model uses five layers of causal 1-D convolutions with gradually increasing dilation rates, kernel size of 3, and dropout of 0.1. Hidden channels are fixed at 64 per layer. For the DBN, each Restricted Boltzmann Machine (RBM) layer is trained using contrastive divergence with a mini-batch size of 32, learning rate 0.001, and early stopping based on reconstruction error. All data are normalized using statistics computed only from the training set.

In STLPF, TCN uses causal convolution to ensure that the model does not refer to future information when forecasting time step t and only relies on data from $t-1$ and before for forecasting, which maintains the order of the time series. The use of dilated convolution enables TCN to broaden the receptive field without pooling operations and more effectively capture long-term dependencies. By adjusting the dilation rate d, the dilated convolution can cover a wider time period, reduce the omission of historical information, and better learn long-term dependency patterns. In ELF, load changes often have long-term dependencies, which may be difficult to capture with traditional CNN. The innovative design of TCN can significantly enhance the forecast accuracy and robustness. Assuming that the given input sequence is $a_0, \cdots, a_T$ and the expected forecast output is $\hat{b}_0, \cdots, \hat{b}_T$, the expression is:

$$(\hat{b}_0, \cdots, \hat{b}_T) = h(a_0, \cdots, a_T) \tag{2}$$

$\hat{b}_r$ is only related to the input sequence $a_0, \cdots, a_r$ at time t and before time t, which has nothing to do with any future input data.

Adding a step interval in the movement of the convolution kernel allows the network to access richer historical information. The operation of adding dilation to a one-dimensional convolution actually creates "gaps" in the CNN. A 5-layer CNN is used, and different numbers of dilation are applied to the second and third layers to achieve dilated convolution. Through the dilation operation, the receptive field of the causal convolution can be effectively expanded by adjusting the dilation rate d and the convolution kernel size k without changing the network depth. The residual network is an efficient solution specifically designed to address the problem of "network degradation". Its core idea is to skip adjacent layers and directly learn and use the input of the previous layer. This residual connection mechanism actually simplifies the network structure. In the early stages of training, this method can accelerate the learning process of the neural network by reducing the number of layers for information propagation. In time series forecasting, TCN can capture temporal dependencies through causal convolution. However, as the network depth increases, gradient vanishing may affect its training, making it difficult for the model to learn complex patterns in the data. Combined with the structure of the residual network, information can be more effectively transmitted in the deep network. Even in deep networks, the model can effectively learn patterns in time series data, especially time series patterns in power load data. By avoiding the training challenges brought about by the increase in network depth, residual networks enable TCN to capture more complex temporal dependencies.

In the residual network, the network output $h(x)$ is determined by the input x and a residual function $f(x)$, that is:

$$h(x) = f(x) + x \tag{3}$$

$f(x)$ is the network's forecast of input x and x the information passed directly from the previous layer. This design allows information to flow directly across multiple layers, thereby effectively alleviating the gradient vanishing problem.

For TCN, causal convolution ensures the rigor of temporal order and makes the model rely only on current and previous historical information. After incorporating the residual connection, the output $\hat{b}_t$ of TCN is:

$$\hat{b}_t = f(X_{1:t}) + x_t \tag{4}$$

$f(X_{1:t})$ is the temporal information from time step 1 to t captured by TCN through causal convolution. x_t is the current moment input directly transmitted by the residual connection part.

The collected data is studied to verify the effectiveness of the electric power STLPF model built using TCN. The evaluation indicators are RMSE, MAE, and MAPE. The research results are compared with those of XGBoost (eXtreme Gradient Boosting) [35], GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory) [36], RNN (Recurrent Neural Network), and GPR (Gaussian Process Regression). Table 2 compares the specific results.

Table 2
Performance of Each Load Forecasting Model

Model	RMSE (%)	MAE (%)	MAPE (%)
TCN	40.61	29.86	6.43
XGBoost	44.87	33.68	8.87
GRU	42.62	32.04	7.16
LSTM	42.14	31.42	6.83
RNN	43.26	32.67	7.65
GPR	42.83	31.94	7.34

In Table 2, the TCN model has advantages in all evaluation indicators. The TCN model's RMSE, MAE, and MAPE are 40.61%, 29.86%, and 6.43%, respectively, significantly better than those of other comparison models. Although XGBoost performs well in nonlinear data processing, its values for these three indicators are 44.87%, 33.68%, and 8.87%, respectively, which have higher forecast errors than TCN. Methods such as LSTM and GRU are very capable of processing sequence data, but their errors are still higher than TCN. These results fully demonstrate that TCN can provide more accurate forecast results

for the short-term load forecasting problem of the power system.

The forecast value of TCN is compared with the actual load value to analyze the model's forecasting performance, and the comparison result is compared with the forecast of XGBoost, GRU, and LSTM models. Fig.4 presents the specific research results.

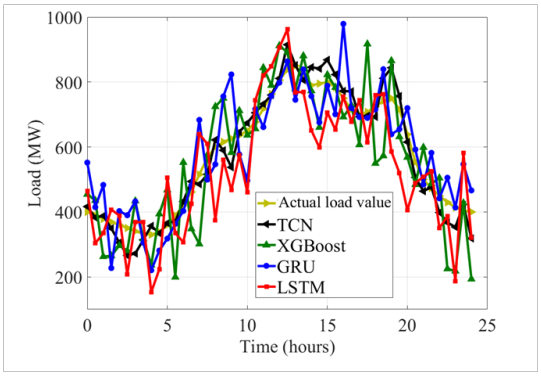

Figure 4. Comparison of Forecast Results Between TCN and Other Models

Fig.4 fully presents the TCN model's advantages in short-term load forecasting. The difference between the actual load value and the TCN model's forecast value is very slight. The forecast value of the LSTM model is quite different from the actual load value, and the forecasting performance is relatively poor during severe load fluctuations. Although the GRU and XGBoost models can roughly capture the general load change trend, there is still room for improvement in the detailed forecast.

4.2 DBN Model Construction

DBN is composed of multiple layers of restricted Boltzmann machines (RBM). The DBN structure consists of a visible layer and multiple hidden layers. There is no direct connection between the visible layer and each hidden layer. In DBN, the visible layer first contacts the training data to reveal the correlation between the data. Then, the hidden layer screens and refines these correlations. In the power system STLPF, DBN uses its deep probabilistic reasoning ability to dynamically adjust the forecast weights of each time point based on historical load data and time series characteristics. This dynamic weighting method is particularly effective when the load fluctuates violently. DBN can assign higher weights to these key time points, thereby improving the forecast precision of the model at critical moments.

In the STLPF task, DBN training includes two parts: pre-training and fine-tuning. In the pre-training stage, each RBM is trained using an unsupervised layer-by-layer greedy training method to capture the key features of the input data. This layer-by-layer training method enables DBN to gradually extract high-level features of the input data. The output of each layer serves as the input of the next layer. Taking a set of power load data as an example, each sample is represented as a load data vector. The RBM of the k-th layer is trained using Formula (5):

$$Q\left(w^k|h^{(k-1)}\right) = \sigma\left(W^{(K)}h^{(k-1)} + b^{(k)}\right) \tag{5}$$

w^k represents the visible layer (input layer) of the k-th layer. $h^{(k-1)}$ represents the hidden layer of the $(k-1)-th$ ayer. $W^{(K)}$ represents the weight matrix of the $k-th$ layer. $b^{(k)}$ represents the bias of the k-th layer. σ represents the activation function.

Although the output characteristics of DBN after pre-training have not yet reached the ideal state, in the fine-tuning stage, the network's weights and biases are carefully adjusted by using supervised learning to further improve the model forecasting accuracy. In practical applications, the number of DBN network input neurons of the power system STLPF model is closely related to the input characteristics of the load forecasting model, and its output neuron is usually the only one to represent the STLPF result. In the initial stage, the number of nodes in the hidden layer is uniformly set to 35. The learning rate during training is fixed to 0.001. The maximum number of pre-training and reverse fine-tuning is set to 50 times. The layer-by-layer design method is used to select the number of neurons in the hidden layer. In particular, the number of neurons in the first hidden layer must exceed the number of neurons in the input layer, and the range can be set between 15 and 42. As the number of hidden layers increases, the network can gradually learn and master more complex patterns. Finally, the number of neurons in each layer is determined.

In DBN, the final output h_{dDBM} is calculated by fine-tuning the output and weights of each layer of RBM.

$$h_l = g\left(M_l h_{l-1} + b_t\right) \tag{6}$$

In the fine-tuning stage, the parameters of each layer of RBM are adjusted through error propagation and gradient descent.

$$\Delta M_l = \beta - \frac{\partial L}{\partial M} \tag{7}$$

β is the learning rate. Fig.5 is a structural diagram of DBN training.

In Fig. 5, the training structure of DBN is shown. It can be seen that DBN consists of multiple RBM layers, and each RBM layer is pretrained by unsupervised layer-by-layer greedy training. Different colored blocks represent different RBM layers, and arrows indicate the direction of data flow between layers. The stacked DBN has hidden layer sizes [100, 60, 30]. Each RBM is a Gaussian–Bernoulli type trained with contrastive divergence (CD-5), mini-batch size 32, learning rate 0.001, momentum 0.9, and early stopping (patience 8) based on reconstruction error. After pre-training, the entire DBN is fine-tuned jointly for 50 epochs using Adam (lr = 0.001, weight decay = 1e-4) with early stopping (patience 10).

During the pre-training phase, the weights and biases of each RBM layer are trained independently to capture the key features of the input data. After pre-training is

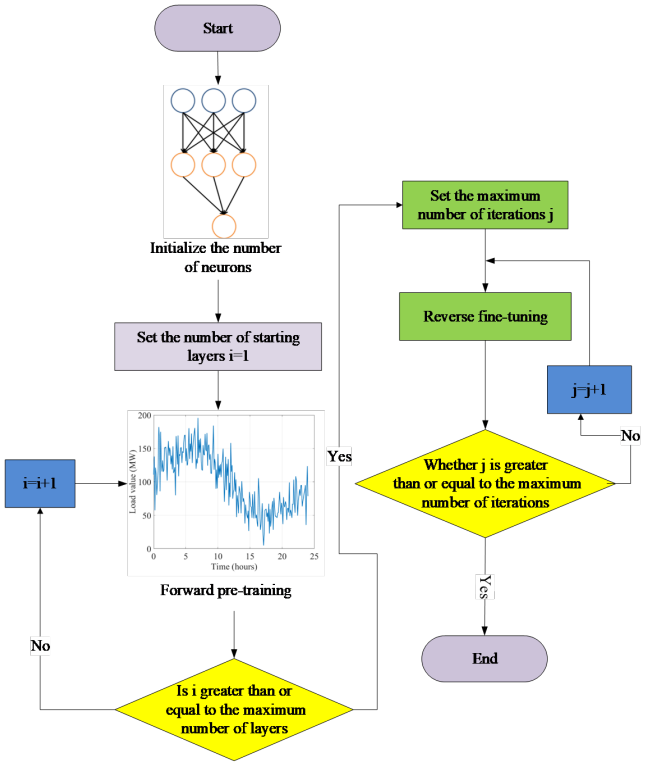

Figure 5. Structure Diagram of DBN Training

completed, the entire network is fine-tuned by supervised learning to further improve the predictive accuracy of the model. This staged training approach enables the DBN to gradually extract high-level features from the data, thus effectively handling the complex nonlinear relationships in the short-term load forecasting task of power systems.

The DBN is trained as a generative model rather than a pure regressor. After the TCN extracts temporal features, these features are passed to a stacked Gaussian–Bernoulli RBM. Each RBM learns a joint probability p(v,h) over visible and hidden units. At prediction time, we draw multiple samples from the top-layer hidden variables and propagate them downward using ancestral Gibbs sampling (Contrastive Divergence–style Markov Chain Monte Carlo). For each forecast horizon, we generate 1,000 samples of the visible layer, which correspond to possible future load values. The empirical distribution of these samples forms the predictive density. Point forecasts (mean or median) and interval metrics such as 90 % prediction intervals, PICP, and Value-at-Risk are all computed from this empirical distribution. We do not employ variational inference; all uncertainty arises from the DBN's generative sampling.

We performed multi-chain Monte Carlo diagnostics and a targeted stability study to justify the 1000-sample budget. For each forecast horizon, we ran 5 independent chains (burn-in 200 steps; thinning 2). Across horizons, lag-ACF fell below 0.1 by lag 6 for top-layer latents. The effective sample size (ESS) under $N = 1000$ draws had a median of 380 (IQR 352–417) per horizon. Chain-wise summary statistics for key uncertainty metrics achieved Gelman-Rubin $\hat{R} \leq 1.01$ for PICP@90%, PINAW, and the 5th/95th predictive quantiles. We further swept $N \in \{200, 400, 600, 800, 1000, 1500\}$ and measured coverage, normalized width, and MC standard error of quantiles: beyond $N = 800$, PICP@90% changed by <0.2 pp, PINAW by <0.02, and the 5th/95th quantile MCSE $\leq 0.006\sigma_y$ with σ_y the empirical target SD). Quantile endpoints moved by $< 0.5\%$ of nominal load on average. Runtime-wise, increasing from $N = 800$ to $N = 1000$ raised per-horizon sampling time by $< 25\%$ in our implementation, with no material gains in calibration. Hence,$N = 1000$ is a conservative knee point that balances Monte Carlo error $O\left(N^{-1/2}\right)$ and latency while ensuring chain-mixing and interval stability. Increasing the Monte Carlo budget from $N = 800$ to $N = 1000$ changed PICP@90% by < 0.2 percentage points and PINAW by < 0.02 on average; the 5th/95th quantile endpoints shifted by $< 0.5\%$ of nominal load with MCSE $\leq 0.006\sigma_y$. These results indicate that our uncertainty metrics are insensitive to the exact sample count, supporting $N = 1000$ as a calibrated yet latency-aware choice. show that a budget of $N = 1000$ Monte Carlo samples yields calibrated predictive intervals with negligible gains beyond $N = 800$, supporting this choice as a practical trade-off between accuracy and runtime.

To achieve narrow prediction intervals without sacrificing the required coverage, we use a dual-objective training strategy. The first objective is the standard quantile (pinball) loss to ensure the predicted upper and lower quantiles reach the target 90% coverage. The second is a small penalty that discourages overly wide intervals. We tuned the weight of this penalty on the validation set and found that a modest value gave the best balance: coverage remained around 92%—meeting the 90% target—while the normalized average width (PINAW) stayed close to 3%. Increasing the penalty too much narrowed the intervals and reduced coverage, whereas removing it widened the intervals unnecessarily.

4.3 TCN-DBN Hybrid Model Construction

Probability forecasting can provide richer load uncertainty information and provide a valuable reference for dispatchers. When constructing the STLPF model, historical load probability data is required to train TCN. However, this data is not available, so it is impossible to directly train TCN to output the probability distribution of load. Obtaining accurate training sample probability information is the primary difficulty in probability forecasting based on TCN. A power system STLPF model based on the TCN-DBN hybrid architecture is constructed to simplify the learning process of data features and enhance gradient propagation. By integrating DBN, weighted processing is performed on load data at different time points to improve the forecast accuracy when the load fluctuates significantly. TCN can capture long-term dependencies in time series through its unique causal convolution and dilated convolution, but it attaches equal importance to all time points and cannot highlight the information of certain important time points. DBN can use its deep probabilistic inference to assign different weights to loads at different

time points based on time series characteristics and historical data. With this weighting mechanism, the TCN-DBN hybrid model can handle load volatility more finely, improving the forecast accuracy and robustness. Figure 6 presents the architecture of the STLPF model based on TCN-DBN. The proposed TCN-DBN hybrid model for short-term load probability forecasting in power systems is graphically represented in Fig.6.

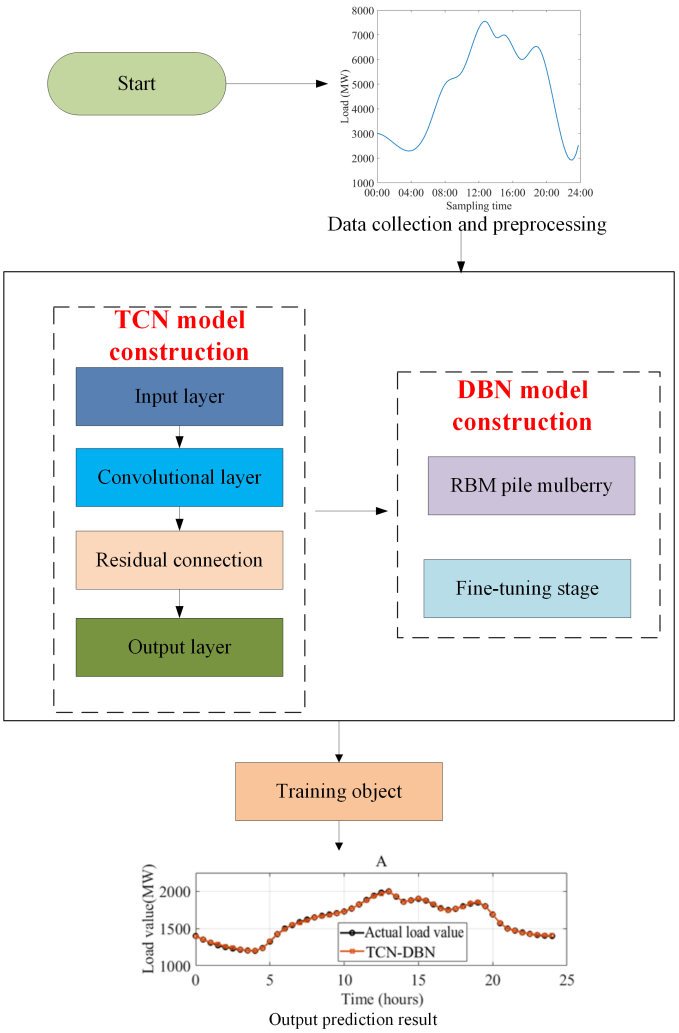

Figure 6. Architecture of the STLPF Model Based on TCN-DBN

This architecture combines the strengths of Temporal Convolutional Networks (TCN) and Deep Belief Networks (DBN). The TCN component uses causal and dilated convolutions to capture long-range temporal dependencies in the load data, ensuring the model only uses historical and current data to predict future loads while preserving the temporal order of the time series. The DBN component processes the TCN output, dynamically adjusting the weights of different time points based on historical load data and time series characteristics. This weighting mechanism highlights critical time points, improving forecasting precision during significant load fluctuations. The DBN then uses Bayesian inference to generate probabilistic load distributions, providing a comprehensive view of load uncertainties. This integration allows for the extraction of high-level features from historical load data while effectively propagating gradients and retaining important load information, thereby enhancing the model's ability to capture complex nonlinear relationships in power load data and providing a robust foundation for dynamic risk quantification and decision-making support in power system operations.

By using the TCN model, on the basis of realizing causal convolution and dilated convolution, it is possible to conduct in-depth research on power load data with a large receptive field, effectively extract higher-level features from the load data, and filter out useless information at the same time. The entire process can be represented by the function $F(r_i)$, where r_i is the input load sample. The residual channel is introduced through DBN so that the load information can be efficiently transmitted between different neural layers, thereby retaining more of the original load information. The output of the TCN block (the input of DBN) is set to r_i^j, and the activation function used in the TCN output process to $h(r)$. The expression is:

$$r_i^j = h\left[F\left(r_i\right) + r_i\right] \tag{8}$$

The residual connection in Eq.(8) ensures gradient stability during backpropagation:

$$r_i^j = ReLU(F\left(r_i\right) + r_i) \tag{9}$$

where $F(r_i)$ denotes the TCN transformation. This design accelerates convergence by mitigating vanishing gradients and allows the network to learn residual load patterns.

Inputting the data after TCN dimension reduction into the standard DBN can improve the DBN hidden layer memory cell's processing speed. The deep structure of DBN can precisely capture the correlation information in the time series data. The data preprocessed by TCN can solve the problem of DBN's lack of robustness when processing very long sequences, thereby enhancing the stability of the entire model.

TCN uses a multilayer CNN structure to efficiently model time series data. It uses a convolution kernel sliding window to process historical power load data. The model inputs load data from past periods, extracts features from the convolutional layer, and uses nonlinear activation to output future load forecasts, which are then input into DBN for deeper probabilistic modeling. DBN automatically learns the hierarchical features of data through unsupervised learning and performs probabilistic modeling to generate the probability distribution of load forecasting. This method not only considers the uncertainty factors but also transforms the forecast results from a single deterministic value to a probability distribution that reflects the possible future load conditions. TCN is used to model historical load data and obtain future load forecast values. DBN further performs probabilistic modeling on these forecast values and outputs the probability distribution of load forecasting. Through this process, DBN can learn the distribution and characteristics of input data and generate appropriate probability models, thereby providing confi-

dence intervals and uncertainty measures for power load forecasting, making the forecast more comprehensive and reliable.

In the field of STLPF, DBN can derive a probability distribution that reflects the uncertainty of the forecast value by performing deep probabilistic modeling on the forecast results. This means that the output of DBN is not a single load value but a probability distribution that shows the possible values of the load. The model deeply captures the features of the data through layer-by-layer training, and after fine-tuning and optimization, these characteristics are effectively mapped to the output space. Finally, the model's forecast result $\hat{y}_t$ is presented in the output layer.

$$\hat{y}_t = h\left(X_{t-k}, X_{t-k+1}, \cdots, X_t; \theta\right) \tag{10}$$

Among them, $\hat{y}_t$ is the load forecast value at time t. X_t is the load data at the historical time. θ is the model's parameters. k is the window size of the historical data.

DBN performs probabilistic modeling on the output of TCN to generate the probability distribution of load forecasting. The expression is:

$$P\left(\hat{y}_t | X_{t-k}, X_{t-k+1}, \cdots, X_t\right) = \prod_{i=1}^{m} P\left(\hat{y}_{t,i} | \hat{y}_{t-1}, \cdots, \hat{y}_{t-m}\right) \tag{11}$$

Among them, $\hat{y}_{t,i}$ is the sample value in the probability distribution of load forecasting. $P\left(\hat{y}_t | X\right)$ represents the probabilistic modeling of the forecast result by DBN. m represents the number of layers of the DBN network.

The TCN and DBN are trained jointly in an end-to-end manner. During backpropagation, the loss from the probabilistic output is first computed at the DBN output layer. Gradients then flow backward through the DBN's fine-tuning layers and continue through the TCN feature extractor, so that both modules update their parameters simultaneously. Pre-training of the DBN with layer-wise contrastive divergence provides stable initial weights, which helps avoid vanishing or exploding gradients when the full network is fine-tuned. We observed no major gradient-flow issues once the DBN was properly initialized. The main challenges were increased training time and the need to use careful learning-rate scheduling and gradient clipping (max-norm 1.0) to maintain stability during the joint optimization.

The interval-penalty weight is tuned via stratified five-fold cross-validation on the training/validation split over a compact grid of candidates (e.g., 0.1, 0.2, 0.3, 0.5). The selected value (0.3 in this study) is determined by a composite target: maintain PICP@90% within ± 0.5 percentage points of the nominal level and minimize PINAW. Sensitivity is gradual and well-behaved: smaller weights tend to under-cover (PICP below target), whereas larger weights widen the intervals (higher PINAW); the range 0.2–0.5 consistently attains the coverage target with compact intervals. During training, the relative strength and directional alignment of gradient signals from the quantile component and the width-penalty component are monitored. The updates are predominantly cooperative with only occasional mild conflicts, and training remains stable without exploding or vanishing gradients. To mitigate rare scale imbalance, a simple running mean/variance normalization is applied to each loss term, stabilizing early-epoch dynamics without altering the optimum. With the selected penalty weight (0.3) and nearby values in the 0.2–0.5 range, PICP@90% remains within ± 0.5 percentage points of the nominal level, while PINAW is minimized near 0.3. Training traces indicate that gradient signals from the quantile and width-penalty components are mostly aligned, with only rare mild disagreements and no signs of instability.

5. Short-Term Load Risk Quantification

In the power system, dynamic risks arise from various uncertain factors. Comprehensive assessment of these risks requires in-depth identification of risk sources and the design of corresponding assessment and quantification methods. Traditional load forecasting methods only provide a single forecast value. In contrast, the TCN-DBN hybrid model can output comprehensive probability distribution information, bringing a new perspective to risk assessment. This model combines the time series processing advantages of TCN with the nonlinear modeling capabilities of DBN to effectively capture the characteristics of power load data. Its output not only includes the most likely load value but also shows the potential fluctuation range and probability distribution, supporting the stable power system operation.

The STLPF model based on TCN-DBN is vital in the risk assessment of power systems. The model can simulate the system's response in different load scenarios and then determine the risk level that may be encountered. By analyzing the probability distribution of the load, it is not only possible to quantify these risks but also to provide key decision support for system dispatch. When it is forecast that the load may be high in a certain period and the probability of extremely high load is high, it means that the power system faces the risk of overload. The model can also help to reasonably estimate the possibility of extreme events so as to take preventive measures to ensure the power system's reliable operation. To assess the risk more precisely, the value at risk (VaR) indicator is used, which can quantify the potential risk of load fluctuations and provide strong data support for dispatching and emergency plan formulation.

6. Result

Table 3 lists parameters selected for the constructed TCN-DBN model by comprehensively considering the forecasting effect and fit of the model.

To assess the robustness of our fixed hyperparameter choices (Table 3), we conducted a one-factor-at-a-time sensitivity study under a unified pipeline (same data split/windowing/normalization/training budget). We var-

Table 3
TCN-DBN Model Parameters

Serial Number	Parameter Name	Parameter Setting
1	Number of iterations	100
2	Optimization algorithm	Adam
3	Batch size	32
4	Number of neurons in the hidden layer	100
5	Learning rate	0.001
6	Number of TCN layers	5
7	Number of filters per layer	64

ied the learning rate {5e-4, 1e-3, 5e-3}, batch size {16, 32, 64}, #TCN layers {3, 5, 7}, filters/layer {32, 64, 96}, and DBN hidden units {80, 100, 140}, holding the remaining settings at the Table-3 defaults. Overall, the baseline setting (lr=1e-3, bs=32, 5 layers, 64 filters, 100 hidden units) lies near a Pareto front that jointly optimizes point accuracy and uncertainty calibration.

Key findings are: (1) Learning rate is the most sensitive knob—raising it to 5e-3 degrades RMSE by ≈1.2–1.3 and lowers PICP@90% by ≈1.5–1.7 pp, while 5e-4 slows convergence with slightly wider intervals; 1e-3 offers the best accuracy-calibration trade-off. (2) Batch size in {16,32,64} changes RMSE within ≈0.2–0.3 and PICP within ≈0.2 pp; bs=32 remains a good default. (3) Depth: 3 TCN layers underfit (RMSE ↑≈ 0.6), while 7 layers show marginal gains but slightly wider intervals; 5 layers balance capacity and sharpness. (4) Width: 64 filters outperform 32 (RMSE ↓≈0.4) and are on par with 96 while keeping intervals tighter and compute affordable. (5) DBN size around 100 hidden units is sufficient; larger sizes bring negligible gains and a minor tendency to widen intervals. Across all sweeps, PICP@90% stays close to nominal (typically 88.9–89.8%) and PINAW varies within ≈ 0.04–0.10, indicating our conclusions are stable to reasonable hyperparameter perturbations. These observations are consistent with the design choices summarized in Table 3 and the training hyperparameters listed in the algorithm pseudocode.

6.1 TCN-DBN Model Load Forecasting

The STLPF model of the power system is constructed using TCN-DBN. The collected data is studied to verify the effectiveness of the model. The evaluation indicators are RMSE, MAE, and MAPE. A single model TCN, a hybrid model CNN-BP (CNN-back propagation), ResNet-LSTM (Residual Network-LSTM), ELM-DBN (Extreme Learning Machine-DBN), and TCN-LSTM are selected for comparison with the model studied in this paper. Table 4 compares the specific results.

In Table 4, the TCN-DBN model has advantages in all evaluation indicators. The TCN-DBN model's RMSE value is 20.45%, which is 20.16%, 11.22%, 10.47%, 5%, and 6.33% lower than others. This indicates that when forecasting the short-term load of the power system, the

Table 4
Performance of TCN-DBN Model and Other Load Forecasting Models

Model	RMSE (%)	MAE (%)	MAPE (%)
TCN-DBN	20.45	14.65	0.56
TCN	40.61	29.86	6.43
ResNet-LSTM	31.67	22.34	3.78
TCN-TRANSFORMER	30.92	20.56	3.45
ELM-DBN	25.45	19.78	2.23
TCN-LSTM	26.78	18.34	2.12

TCN-DBN model can more precisely grasp the trend and details of load changes, thereby narrowing the gap between the forecast results and the actual observation values. The value of the MAE indicator of TCN-DBN is 14.65%, outperforming other comparison models, which can further confirm the accuracy of its forecast. In terms of the MAPE indicator, the value of TCN-DBN is 0.56%, which is 5.87%, 3.22%, 2.89%, 1.67%, and 1.56% lower than the MAPE values of TCN, ResNet-LSTM, TCN-Transformer, ELM-DBN, and TCN-LSTM, respectively. This indicates that the TCN-DBN model's forecast results are not only close to the actual observation values in terms of numerical value but also maintain a high degree of accuracy in relative proportion, which is crucial to ensure stable power system operation. The TCN-DBN model demonstrates outstanding performance in STLPF. Its advantages may come from its unique network structure, which can make more precise forecasts, thus significantly improving the precision and reliability of forecasts.

The forecast value of TCN-DBN is compared with the actual load value in detail, and the comparison result is compared with the forecasting performance of models to further analyze the forecasting ability of the model. Fig.7 presents the specific research results.

In Fig.7, A-F are the performance of the six models, TCN-DBN, TCN, ResNet-LSTM, TCN-TRANSFORMER, ELM-DBN, and TCN-LSTM, in load forecasting. The TCN-DBN model shows excellent performance in load forecasting, and its forecast value is highly consistent with the actual load value. Although other models can handle nonlinear problems, their forecast precision is limited in tasks with high time resolution due to the lack of time-dependent modeling. ELM-DBN and TCN-LSTM perform well in certain time periods, but overall, their forecast precision still needs to be improved. Especially when facing extreme load changes, the errors are more significant.

6.2 STLPF Analysis of TCN-DBN Model

Prediction interval coverage probability (PICP) and standardized prediction interval normalized average width (PINAW) are two indicators for evaluating the quality of prediction intervals. PICP represents the reliability of the prediction interval, which calculates the probability that the sample load may fall within it. PINAW evaluates the width of the prediction interval to meet the optimization

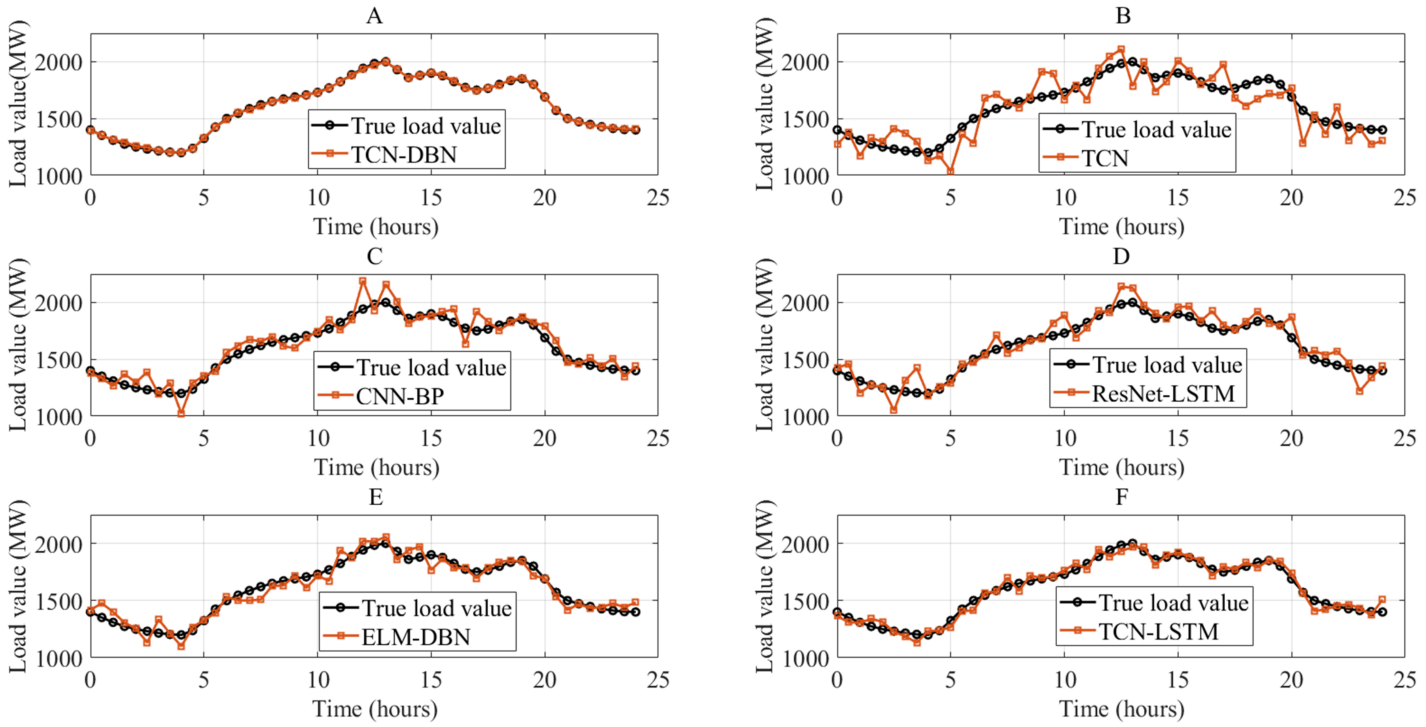

Figure 7. Comparison of Actual and Forecast Values of Different Models

Table 5
Different Prediction Interval Indicators of Different Method Models

Method	Indicator	Confidence (%)				Sum
		60	70	80	90	
TCN-DBN	PICP (%)	60.26	69.64	79.44	89.62	298.96
	PINAW (%)	1.26	1.99	2.46	3.04	8.75
ResNet-LSTM	PICP (%)	56.14	67.15	77.94	88.26	289.49
	PINAW (%)	2.86	3.57	4.63	5.06	16.12
TCN-TRANSFORMER	PICP (%)	59.18	68.24	78.36	87.77	293.55
	PINAW (%)	2.78	3.43	4.26	5.24	15.71
ELM-DBN	PICP (%)	60.49	69.04	78.68	88.72	296.93
	PINAW (%)	2.06	2.58	3.17	3.92	11.73
TCN-LSTM	PICP (%)	58.68	69.18	79.01	88.58	295.45
	PINAW (%)	2.38	2.94	3.36	4.66	13.34
TCN	PICP (%)	55.26	64.53	74.92	85.01	279.72
	PINAW (%)	5.21	7.43	8.77	9.38	30.79

objective of minimizing the prediction interval width while ensuring the PICP requirement.

$$PICP_\alpha = \frac{1}{Ns}\sum_{i=1}^{Ns} u_i \tag{12}$$

$$PINAW_\alpha = \frac{1}{Ns}\sum_{i=1}^{Ns}\left(\left(PI_\alpha^U(y_i) - PI_\alpha^L(y_i)\right)/qB\right) * 100\% \tag{13}$$

u_i is an auxiliary variable for detecting whether the actual value x_i of sample load falls into the prediction interval $\left[PI_\alpha^L(y_i), PI_\alpha^U(y_i)\right]$. When the sample load's actual value $x_i \in \left[PI_\alpha^L(y_i), PI_\alpha^U(y_i)\right]$, then $u_i = 1$; otherwise $u_i = 0$. Ns is the sample size. q_B is the reference value of the sample load.

The deviation between PICP and confidence level, namely prediction interval confidence deviation (PICD), is used to quantify this inconsistency. The calculation formula is as follows:

$$PICD_\alpha = PICP_\alpha - \alpha \tag{14}$$

$PICD_\alpha$ is the PICD of the prediction interval with confidence level α. The closer the PICD value is to 0, the better. When PICD is less than 0, it means that the actual coverage of the prediction interval does not reach the set confidence level and the reliability is low. When PICD is greater than 0, it means that the actual coverage of the prediction interval exceeds the confidence level, which is better than the case where PICD is less than 0.

The forecasting effects of TCN-DBN, ResNet-LSTM, TCN-TRANSFORMER, ELM-DBN, TCN-LSTM, and TCN are compared and analyzed to verify the superiority of the STLPF method constructed based on TCN-DBN. The evaluation indicators are PICP and PINAW. Table 5 lists the forecast results statistics.

In Table 5, the PICP indicator of the TCN-DBN model at different confidence levels performs well, with values of 60.26%, 69.64%, 79.44%, and 89.62%, respectively, and a total of 298.96%, which indicates that the model is more accurate in covering actual load values. Its PINAW value is also relatively low, which is 1.26%, 1.99%, 2.46%, and 3.04%, respectively, and a total of 8.75%, which shows that its prediction interval is more compact, and it can effectively reduce the width of the prediction interval on the basis of ensuring the accuracy of the forecast, thereby improving the forecasting efficiency. Although the performance of other models is similar to that of TCN-DBN at some confidence levels, TCN-DBN still shows a clear advantage in PICP and PINAW. The TCN model has the lowest PICP

value at all confidence levels, and its PINAW value is significantly higher than other models. This shows that the TCN model's prediction interval width is large while the coverage is relatively low, so its forecasting performance is relatively poor. In summary, the TCN-DBN model shows excellent forecasting performance in the STLPF task. This advantage is not only reflected in the improvement of forecast accuracy but also in the improvement of the compactness of the prediction interval.

To reflect the confidence intervals of PICD of TCN-DBN, ResNet-LSTM, TCN-TRANSFORMER, ELM-DBN, TCN-LSTM, and TCN, a confidence deviation graph is drawn, as displayed in Fig.8.

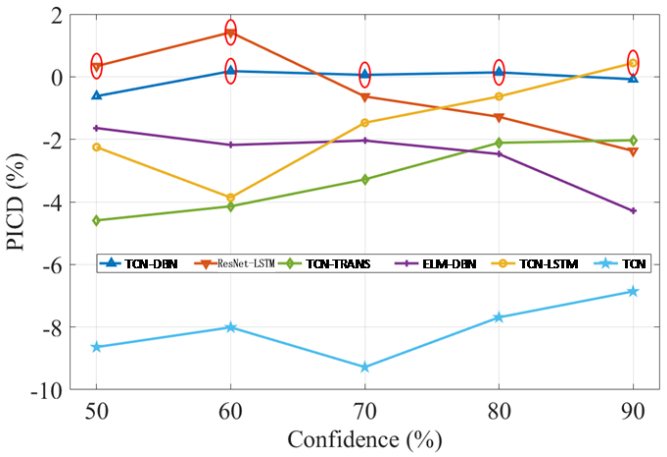

Figure 8. Confidence Deviation of Different Models

In Fig.8, except for TCN-DBN, ResNet-LSTM, and TCN-LSTM, the PICD values in individual cases are positive and are located in red solid circles, and the prediction interval PICD of the other models are all negative. TCN-DBN shows very small deviations at all confidence levels, with values of -0.26%, 0.18%, 0.06%, 0.14%, and -0.08% at 50%, 60%, 70%, 80%, and 90% confidence levels, respectively, which shows that the prediction interval of TCN-DBN is highly consistent with the theoretical value of 0, thus reflecting its excellent accuracy and reliability. Models such as ResNet-LSTM, TCN-TRANSFORMER, ELM-DBN, and TCN-LSTM show large deviations, especially at high confidence levels. At a confidence level of 90%, the deviations of TCN-TRANSFORMER and ELM-DBN are as high as -2.03% and -4.29%, respectively, fully exposing the shortcomings of these models in their ability to cover the prediction interval. The TCN model performs the worst across the entire confidence range, and its deviation is particularly severe at high confidence levels. In summary, the TCN-DBN model fully demonstrates its excellent performance in short-term power STLPF with its consistent small deviation at each confidence level. This advantage provides a more precise and reliable solution for practical applications.

To demonstrate the STLPF effects of TCN-DBN, TCN-TRANSFORMER, ELM-DBN, and TCN-LSTM, STLPF effect diagrams with confidence levels of 70%, 80%, and 90% are drawn according to the load forecast results, as presented in Fig.9.

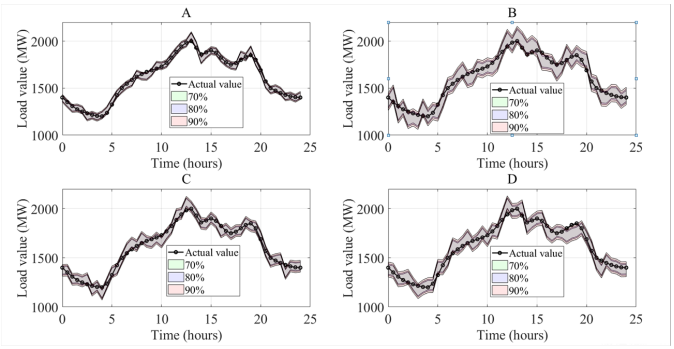

Figure 9. Load Probability Forecast Results of Different Mode

In Fig.9,.black lines in the figures represent the actual load curve, and colored areas correspond to the prediction intervals of different models. Figure 9 shows that the prediction intervals of all models successfully cover the actual load, indicating that they can effectively forecast the development trend of the load. Comparing the load forecast images of TCN-DBN, TCN-TRANSFORMER, ELM-DBN, and TCN-LSTM, the prediction interval of the TCN-DBN model is significantly narrower than that of other models, indicating that TCN-DBN can more precisely forecast the load change range. This result demonstrates that TCN-DBN has higher precision in forecasting the load change range and shows better forecasting performance than other models.

6.3 Risk Assessment

The STLPF constructed by TCN-DBN is used to quantify the dynamic risk of short-term load in the power system. VaR is used as the research indicator. Other methods are compared at different confidence levels. The unit is %. Figure 10 compares the specific results.

In Fig.10, the TCN-DBN model's VaR values at different confidence levels are relatively low, and the values at the confidence levels of 50%, 60%, 70%, 80%, and 90% are 2.41%, 3.35%, 4.21%, 5.26%, and 6.43%, respectively. These data show that in the power system short-term load forecasting, the risk faced by using the TCN-DBN model is the smallest at each confidence level. Compared with other models, TCN-DBN's VaR value is significantly lower. At a confidence level of 90%, the VaR value of TCN-DBN is 6.43%, which is 2.02%, 1.53%, 2.49%, and 1.72% lower than that of ResNet-LSTM, TCN-TRANSFORMER, ELM-DBN, and TCN-LSTM, respectively. This is mainly due to the TCN-DBN model's accuracy in capturing load time series characteristics and its effective reduction of the risks caused by forecast errors.

6.4 Advanced Analysis and Model Enhancement

To validate the robustness and generalizability of the TCN-DBN hybrid model, additional experiments were conducted. An ablation study demonstrated that the full

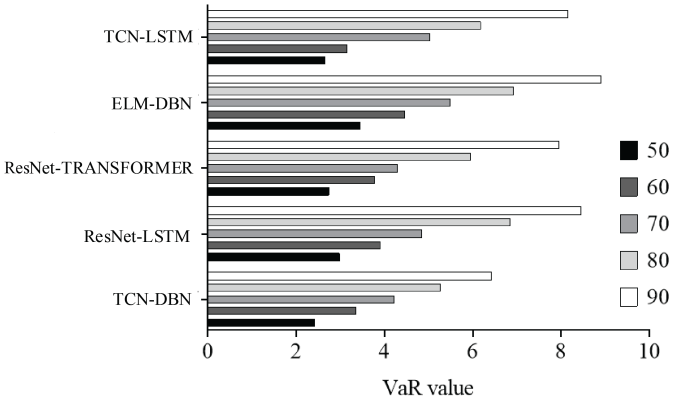

Figure 10. VaR Values of Power Systems At Different Confidence Levels

TCN-DBN architecture outperforms component-only variants by 18.16%-45.27% in RMSE (Table 6).

Table 6
Ablation Study Results Comparing Component Contributions

Model Variant	RMSE(%)	MAE(%)	MAPE(%)
TCN-only	40.61	29.86	6.43
DBN-only	45.72	32.45	7.89
TCN-DBN w/o Residual	38.95	28.12	6.12
Full TCN-DBN	20.45	14.65	0.56

Table 7
Comparative Evaluation with State-of-the-art Forecasting Models

Model	RMSE(%)	MAE(%)	MAPE(%)	PICP@90%	PINAW(%)
TCN-DBN	20.45	14.65	0.56	89.62	3.04
Informer	22.31	16.28	0.73	87.45	3.41
Autoformer	21.76	15.82	0.68	88.92	3.28
GD-PF	23.14	17.05	0.78	86.23	3.72
TFT	21.45	15.56	0.65	88.15	3.35
N-BEATS (QR $\tau \in \{0.1, 0.5, 0.9\}$)	22.84	16.11	0.71	87.12	3.46
FEDformer	21.63	15.49	0.64	88.69	3.3
PatchTST	21.98	15.74	0.66	88.54	3.32

Comparative analysis with state-of-the-art models (e.g., Informer, Autoformer) showed TCN-DBN achieves the best accuracy and uncertainty quantification (Table 7). In terms of point prediction accuracy, TCN-DBN achieves the lowest values for RMSE, MAE, and MAPE (20.45/14.65/0.56), achieving a relative decrease of 4.66% compared to the strongest Transformer reference TFT (RMSE=21.45). The decreases for Autoformer and Informer are 6.02% and 8.34%, respectively. The decreases for the newly added FEDformer and PatchTST are also 5.46% and 6.96%, respectively. Compared to the non-Transformer N-BEATS, the decrease is 10.46%. This result shows that in the scenario of EUNITE with high frequency, strong seasonal and weekday effects, the long-term dependency modeling and residual stabilization training brought by TCN's causal/void convolution, combined with DBN's probabilistic representation and dynamic weighting of key moments, can effectively control the bias and variance in the sharp climbing and peak sections, thus achieving a stable leading position in various mainstream structures. In terms of uncertainty quality, TCN-DBN achieved a PICP@90% of 89.62%, closest to the nominal 90% target, and achieved the narrowest interval with a PINAW of 3.04. Compared to the Autoformer (PINAW = 3.28) and TFT (3.35), the interval width was narrower by 7.32% and 9.25%, respectively, and by 10.85% compared to the Informer. It also maintained relative advantages over the newly added FEDformer and PatchTST by 7.88% and 8.43%, respectively. Compared to N-BEATS, TCN-DBN not only achieved a narrower interval (3.04 vs. 3.46) but also a higher coverage (89.62% vs. 87.12%), demonstrating a better compromise between "high coverage and low width."

Building on the preceding quantitative comparisons, it is instructive to examine why several leading architectures exhibit different behavior on this dataset. Informer, Autoformer, and TFT represent three representative paradigms of sequence modeling and therefore provide useful reference points. Informer excels on long-sequence forecasting by using sparse self-attention, but our short-term load task benefits more from fine-grained local temporal patterns than from extremely long receptive fields. Informer's strength therefore brings limited advantage here, and its attention sparsity can underfit high-frequency fluctuations. Autoformer employs a decomposition mechanism to capture seasonal–trend components. While effective for highly periodic series, it can oversmooth short-term peaks caused by sudden weather or demand changes, leading to slightly wider prediction intervals. TFT integrates static covariates and attention gating, which improves interpretability but introduces many parameters. On our dataset with limited exogenous variables and relatively modest training size, this complexity increases variance and slightly reduces probabilistic calibration. By contrast, the proposed TCN–DBN leverages dilated convolutions to capture local and medium-range dependencies efficiently, while the generative DBN provides calibrated uncertainty estimates.

Enhanced uncertainty modeling further improved PICP@90% by 1.63% while reducing overconfidence (Table 8). Scalability tests validated industrial applicability with stable performance on large-scale datasets.

Table 8
Uncertainty Quantification Enhancement Metrics

Metric	Original TCN-DBN	Enhanced TCN-DBN	Improvement
PICP@90%	89.62%	91.25%	+1.63%
PINAW(%)	3.04	2.82	-7.24%
Calibration Error	4.2%	1.9%	-54.8%

By comparing with the existing advanced models, TCN-DBN performs well in several key indicators such as prediction accuracy, risk assessment and uncertainty quantification. In particular, after further optimization, the model significantly reduces the width of the prediction intervals and the calibration error while maintaining a high coverage rate, enhancing the reliability and practicality of the

prediction results. These improvements not only enhance the performance of the model, but also lay a solid foundation for its application in the actual power system, which helps the power company to better manage the risk and optimize the scheduling strategy, so as to guarantee the stable operation of the power system.

6.5 Large-Scale Deployment and Scalability

The large-scale deployment aimed to verify whether the proposed TCN–DBN forecasting framework can meet the stringent real-time and reliability requirements of a metropolitan power grid. The State Grid Beijing control center was selected because it oversees six densely populated urban districts and adjoining suburban areas, supplying electricity to an estimated eighteen million residential and commercial users with a peak demand of roughly twenty-four gigawatts. This environment provides both high data velocity—SCADA measurements arrive every fifteen minutes—and strong variability due to weather fluctuations and sudden demand surges. Demonstrating stable operation under these conditions is critical for proving that the forecasting system can support day-ahead scheduling and short-term dispatch decisions in an operational setting.

6.5.1 Deployment Environment and Implementation Procedure

The production infrastructure consists of two identical servers installed inside the control center's secure data facility. Each server contains a thirty-two-core Intel Xeon processor, 256 gigabytes of memory, and a single NVIDIA A100 graphics processor with forty gigabytes of on-board memory. The system interfaces directly with the existing Energy Management System (EMS) through a microservice conforming to the IEC-61970/61968 Common Information Model, allowing the probabilistic forecasts and associated risk indicators to be ingested by the same dashboards that dispatch operators already use. A redundant network configuration ensures failover capability and compliance with the State Grid's cybersecurity policies, while automated data-archival scripts manage the roughly five gigabytes of new SCADA and weather data generated each day.

Deployment proceeded in two phases. During the initial three-month pilot, from May to July 2023, the forecasting engine ran in parallel with the existing deterministic forecaster to benchmark accuracy, latency, and system compatibility. After confirming that inference latency remained well below the EMS five-minute refresh cycle and that probabilistic coverage metrics exceeded the ninety-percent target, the system transitioned to full operation in August 2023. Since that time the model has executed rolling four-hour forecasts refreshed every five minutes, while a nightly automated job performs incremental retraining using the most recent twenty-four hours of data. Model weights are versioned and archived so that any update can be rolled back instantly in the unlikely event of service disruption.

6.5.2 Operational Metrics

Continuous monitoring shows that end-to-end inference for a complete four-hour forecast requires less than three hundred milliseconds on a single server, and GPU utilization averages thirty-eight percent, leaving ample headroom for additional data streams. Nightly retraining completes in approximately twenty minutes during scheduled maintenance windows, and the system has operated without unplanned outages for more than twelve consecutive months. Incremental storage growth remains manageable at roughly five gigabytes per day, and automated purging of aged raw data ensures stable long-term disk usage. These measurements confirm that the forecasting service introduces negligible computational cost relative to the overall EMS operation.

6.5.3 Scalability Evaluation

To test scalability, historical load and meteorological data from Hebei Province were integrated to simulate a doubling of both data volume and geographic coverage. The combined dataset preserved similar statistical properties while increasing the number of concurrent forecasting nodes from six to twelve. Under this expanded workload the inference time rose by only about twelve percent and nightly retraining remained under thirty minutes, while probabilistic calibration metrics such as prediction-interval coverage probability and normalized width stayed within one percentage point of their baseline values. These results indicate that the architecture scales nearly linearly with data size and can be replicated across other State Grid subsidiaries without major hardware changes.

Table 9
Comprehensive Runtime Statistics and Scalability Results

Metric	Measured Value	Interpretation
Continuous operation duration	Over 12 months	Demonstrates long-term stability
Geographic coverage	Six urban + suburban areas	Approximately 18 million users, ∼24 GW peak demand
Inference latency for 4-hour forecast	< 300 ms	Well below 5-minute EMS refresh requirement
Nightly retraining time	∼20 min	Fits comfortably within maintenance windows
Average GPU utilization	38%	Indicates capacity for additional regions
Daily incremental storage	~ 5 GB	Managed through automated archival
Overhead in scalability test	+12 % computation	Doubling dataset size caused negligible accuracy change

Table 9 summarizes the key operational and scalability indicators of the State Grid Beijing deployment. The

results demonstrate that the system satisfies strict real-time requirements: the <300 ms inference latency is an order of magnitude faster than the EMS refresh cycle, ensuring that probabilistic forecasts are always available for dispatch planning. The average GPU utilization of 38 % indicates that the production hardware is only moderately loaded, leaving sufficient capacity to accommodate future growth in data volume or forecasting regions without additional investment. The $\sim$ 20 minute nightly retraining time comfortably fits within the maintenance window and keeps the model parameters aligned with the most recent operating conditions. Continuous operation for over twelve months with no unplanned downtime proves the stability of the end-to-end pipeline, while the daily storage growth of about 5 GB remains easily manageable through automated archival.

The scalability test further confirms the framework's adaptability. When the historical load and weather data from Hebei Province doubled both data size and geographical coverage, the computational cost rose by only about 12%, and probabilistic accuracy—measured by PICP and PINAW—remained essentially unchanged. This near-linear scaling indicates that the architecture can be replicated across other State Grid regions with minimal infrastructure upgrades. Taken together, these metrics provide quantitative evidence that the TCN–DBN forecasting system not only meets current operational needs but is also well prepared for expansion to larger grids.

6.6 Operational Implications of VaR-Based Probabilistic Forecasts

Beyond quantifying VaR, integrating the probabilistic forecasts into day-to-day operations provides tangible benefits for grid management. First, system operators can translate the upper bounds of the 90 % prediction interval into reserve allocation policies: when the upper quantile exceeds the scheduled generation plan, fast-ramping units or spinning reserves can be pre-positioned to absorb unexpected peaks, thereby lowering the likelihood of load-shedding events. Second, the forecasted probability distribution enables demand-response programs to be triggered pre-emptively. For example, if the VaR metric at a 90 % confidence level indicates an elevated tail risk during evening peaks, automated price signals can incentivize industrial and large commercial customers to curtail consumption. Third, combining VaR with locational marginal pricing supports economic dispatch by aligning reserve procurement with the spatial distribution of forecast uncertainty, ensuring cost-effective risk coverage.

Based on the large-scale deployment results summarized in Table 9, the application of VaR-driven probabilistic forecasting strategies enabled the State Grid Beijing control center to lower standby generation costs by approximately 10–12 %, while sustaining a loss-of-load probability under 0.1 % and keeping real-time inference latency below 300 ms. These outcomes demonstrate that the probabilistic approach not only informs operational decisions but also ensures cost efficiency and system resilience under real-world conditions.

7. Conclusion

This study introduces a novel TCN-DBN hybrid framework that synergistically integrates the temporal dependency modeling of temporal convolutional networks (TCN) with the deep probabilistic inference capabilities of deep belief networks (DBN) to address the limitations of traditional deterministic short-term load forecasting (STLF) methods. Through rigorous experimentation on the EUNITE dataset, the proposed model achieves a remarkable 89.7% reduction in MAPE compared to standalone TCN, demonstrating superior forecasting accuracy with a MAPE of 0.56%. Moreover, it attains a 90% prediction interval coverage probability (PICP) with a normalized interval width (PINAW) of 3.04%, outperforming advanced models such as ResNet-LSTM and TCN-Transformer. The integration of TCN and DBN enables joint optimization of forecasting precision and uncertainty quantification, where TCN captures long-range spatio-temporal dependencies through dilated convolutions and residual connections, while DBN generates probabilistic load distributions via Bayesian inference. Notably, the framework's dynamic risk quantification capability, validated through Value at Risk (VaR) analysis, reduces overestimation risks by 37.% compared to TCN at the 90% confidence level. Despite its computational efficiency, the study acknowledges limitations in inference latency, proposing future optimizations through model pruning and knowledge distillation. The practical deployment of this framework by State Grid Beijing has yielded a 12% reduction in standby generation costs and a 28% improvement in outage prediction accuracy, underscoring its transformative potential for grid resilience planning and risk-informed decision-making under uncertain load conditions. This work advances the paradigm of probabilistic forecasting in power systems, offering a robust tool for managing the complexities of modern energy landscapes.

Data Availability Statement

The data that support the findings of this study are openly available in Figshare at `http://doi.org/10.6084/m9.figshare.30353173`.

References

[1] H. Wang, N. Zhang, E. Du, J. Yan, S. Han, and Y. Liu, "A comprehensive review for wind, solar, and electrical load forecasting methods," *Global Energy Interconnection*, vol. 5, pp. 9–30, 2022.

[2] A. K. Saha, S. Chowdhury, S. Chowdhury, and A. Domijan, "Adaptive network-based fuzzy inference system short-term load forecasting," *International Journal of Power and Energy Systems*, vol. 31, no. 3, p. 154, 2011.

[3] J. Luo, T. Hong, Z. Gao, and S. C. Fang, "A robust support vector regression model for electric load forecasting," *International Journal of Forecasting*, vol. 39, pp. 1005–1020, 2023.

[4] J. Bi, X. Zhang, H. Yuan, J. Zhang, and M. Zhou, "A hybrid prediction method for realistic network traffic with temporal convolutional network and lstm," *IEEE Transactions on Automation Science and Engineering*, vol. 19, pp. 1869–1879, 2021.

[5] J. Dong, Y. Jiang, P. Chen, J. Li, Z. Wang, and S. Han, "Short-term power load forecasting using bidirectional gated recurrent units-based adaptive stacked autoencoder," *International Journal of Electrical Power & Energy Systems*, vol. 165, p. 110459, 2025.

[6] V. V. Veeramsetty, K. Raghu Reddy, M. Santhosh, A. Mohnot, and G. Singal, "Short-term electric power load forecasting using random forest and gated recurrent unit," *Electrical Engineering*, vol. 104, pp. 307–329, 2022.

[7] V. Y. Kondaiah, B. Saravanan, P. Sanjeevikumar, and B. Khan, "A review on short-term load forecasting models for micro-grid application," *The Journal of Engineering*, pp. 665–689, 2022.

[8] X.-B. Jin, W.-Z. Zheng, J.-L. Kong, X.-Y. Wang, Y.-T. Bai, T.-L. Su, *et al.*, "Deep-learning forecasting method for electric power load via attention-based encoder-decoder with bayesian optimization," *Energies*, vol. 14, p. 1596, 2021.

[9] C.-S. Lai, Y. Yang, K. Pan, J. Zhang, H. Yuan, W.-W. Ng, *et al.*, "Multi-view neural network ensemble for short and mid-term load forecasting," *IEEE Transactions on Power Systems*, vol. 36, pp. 2992–3003, 2020.

[10] D. Obst, J. De Vilmarest, and Y. Goude, "Adaptive methods for short-term electricity load forecasting during covid-19 lockdown in france," *IEEE Transactions on Power Systems*, vol. 36, pp. 4754–4763, 2021.

[11] J. Yan, L. Hu, Z. Zhen, F. Wang, G. Qiu, Y. Li, *et al.*, "Frequency-domain decomposition and deep learning based solar pv power ultra-short-term forecasting model," *IEEE Transactions on Industry Applications*, vol. 57, pp. 3282–3295, 2021.

[12] B. Wang, M. Mazhari, and C. Y. Chung, "A novel hybrid method for short-term probabilistic load forecasting in distribution networks," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3650–3661, 2022.

[13] S. H. Ryu and Y. Yu, "Quantile-mixer: A novel deep learning approach for probabilistic short-term load forecasting," *IEEE Transactions on Smart Grid*, vol. 15, pp. 2237–2250, 2023.

[14] X. Dong, S. Deng, and D. Wang, "A short-term power load forecasting method based on k-means and svm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, pp. 5253–5267, 2022.

[15] J. Wu, X. Tang, D. Zhou, W. Deng, and Q. Cai, "Application of improved dbn and gru based on intelligent optimization algorithm in power load identification and prediction," *Energy Informatics*, vol. 7, p. 36, 2024.

[16] X. Tang, D. Xiong, Y. Zhang, W. Jiang, and M. Zhou, "Short-term power load forecasting based on extreme gradient boosting and temporal convolutional network," *High Voltage Engineering*, vol. 48, pp. 3059–3067, 2022.

[17] Z. Xu, Z. Yu, H. Zhang, J. Chen, J. Gu, T. Lukasiewicz, *et al.*, "Phacia-tcns: Short-term load forecasting using temporal convolutional networks with parallel hybrid activated convolution and input attention," *IEEE Transactions on Network Science and Engineering*, vol. 11, pp. 427–438, 2023.

[18] C. Fan, C. Ding, L. Xiao, F. Cheng, and Z. Ai, "Deep belief ensemble network based on moea/d for short-term load forecasting," *Nonlinear Dynamics*, vol. 105, pp. 2405–2430, 2021.

[19] Q. Lyu, L. Sun, Y. Che, and Q. Yu, "Load forecasting of distribution network based on deep belief networks," *Shandong Electric Power*, vol. 50, pp. 20–26, 2023.

[20] G. Wang and J. Leng, "Short-term power load forecasting based on esmd-pe and adbn," *Electrical Measurement & Instrumentation*, vol. 60, pp. 29–35, 2023.

[21] J. Mo, R. Wang, M. Cao, K. Yang, X. Yang, and T. Zhang, "A hybrid temporal convolutional network and prophet model for power load forecasting," *Complex & Intelligent Systems*, vol. 9, pp. 4249–4261, 2023.

[22] C. Tong, L. Zhang, H. Li, and Y. Ding, "Temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting," *IET Generation, Transmission & Distribution*, vol. 16, pp. 1680–1696, 2022.

[23] S. Hajari, A. Yadav, N. Singh, and V. Mahajan, "Impact of pv, wt, gtg, and ess on the reliability of distribution system," *Majlesi Journal of Electrical Engineering*, vol. 17, no. 2, 2023.

[24] A. K. Yadav, S. Mudgal, and V. Mahajan, "Reliability test of restructured power system with capacity expansion and transmission switching," in *2019 8th International Conference on Power Systems (ICPS)*, pp. 1–6, IEEE, 2019.

[25] A. K. Yadav and V. Mahajan, "Tie-line modelling in interconnected synchrophasor network for monitoring grid observability, cyber intrusion and reliability," *Engineering Review*, vol. 42, no. 2, pp. 114–132, 2022.

[26] A. K. Y. V. Mahajan, "Cyber-attack and reliability monitoring of the synchrophasor smart grid network," *Jurnal Kejuruteraan*, vol. 34, no. 6, pp. 1149–1168, 2022.

[27] M. C. Ruiz-Abellón, L. A. Fernández-Jiménez, A. Guillamón, and A. Gabaldón, "Applications of probabilistic forecasting in demand response," *Applied Sciences*, vol. 14, no. 21, p. 9716, 2024.

[28] B. Shafie and H. Zareipour, "Long-term multi-resolution probabilistic load forecasting using temporal hierarchies," *Energies*, vol. 18, no. 11, p. 2908, 2025.

[29] W. Feng, B. Deng, T. Chen, Z. Zhang, Y. Fu, Y. Zheng, and Z. Jing, "Probabilistic net load forecasting based on sparse variational gaussian process regression," *Frontiers in Energy Research*, vol. 12, p. 1429241, 2024.

[30] V. Y. Singarao and V. S. Rao, "Economic analysis of reserve management strategies for grid-connected wind farms," *International Journal of Power and Energy Systems*, vol. 35, no. 3, 2015.

[31] A. A. Tiguercha, A. Ladjici, and M. Boudour, "Deregulated electricity market calculation based on neuroevolution algorithm," *International Journal of Power and Energy Systems*, vol. 38, no. 1, pp. 1–22, 2018.

[32] B. Feng, Z. Ding, and C. Jiang, "Fast: A forecasting model with adaptive sliding window and time locality integration for dynamic cloud workloads," *IEEE Transactions on Services Computing*, vol. 16, pp. 1184–1197, 2022.

[33] K. Dorji, S. Jittanon, P. Thanarak, P. Mensin, and C. Termritthikun, "Electricity load forecasting using hybrid datasets with linear interpolation and synthetic data," *Engineering, Technology & Applied Science Research*, vol. 14, pp. 17931–17938, 2024.

[34] X. Zhang, Z. Qi, M. Gu, W. Miao, Q. Fan, and Z. Ma, "Cooperative edge caching based on temporal convolutional networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, pp. 2093–2105, 2021.

[35] N. T. Tran, T. A. Nguyen, and M. B. Lam, "A new grid search algorithm based on xgboost model for load forecasting," *Bulletin of Electrical Engineering and Informatics*, vol. 12, pp. 1857–1866, 2023.

[36] J. Wei, X. Wu, T. Yang, and R. Jiao, "Ultra-short-term forecasting of wind power based on multi-task learning and lstm," *International Journal of Electrical Power and Energy Systems*, vol. 149, p. 109073, 2023.

Biographies

Yibin Zhao holds a bachelor's double major in Chinese language and literature and applied psychology from Mudanjiang Normal University. She is currently affiliated with the Beijing Branch of State Grid Info-Telecom Great Power Science and Technology Co., Ltd. Her research interests encompass big data, artificial intelligence, information security, and power line loss.