Dual-Stream CNN-GRU Model with Spatial and Self-Attention for Power Forecasting in Electrical Automation

Chengjiang Tang
Heilongjiang Vocational College of Agricultural Technology
Jiamusi, Heilongjiang, 154007, China
Email: hnzytcj@126.com

Keywords: Electrical automation, smart grid, power forecasting, hybrid model

Received: February 23, 2025

Accurate power forecasting is critical for electrical automation within smart grids, enabling intelligent energy management to balance demand and supply. We propose a dual-stream hybrid model that processes multivariate time series data from the DKASC (solar generation) and IHEPC (household consumption) datasets through parallel Convolutional Neural Networks (CNNs) with spatial attention and Gated Recurrent Units (GRUs) with self-attention. This unified model captures cross-variable correlations (e.g., weather, power output) and long-term temporal dependencies to predict both generation and consumption, enhancing grid stability. Evaluated using RMSE, MAE, and MSE, our model achieves a 32–93% reduction in RMSE compared to baseline methods like RCC-LSTM and CNN-LSTM hybrids on DKASC and IHEPC datasets. Designed for integration with electrical automation systems, it offers superior accuracy, robustness, and efficiency, advancing smart grid operations.

Povzetek: Članek predstavi dvojni tok CNN-GRU model z vgrajenima prostorsko in samo-pazljivostjo za napovedovanje proizvodnje in porabe elektrike. Model izboljšuje robustnost in učinkovitost pametnih omrežij.

1 Introduction

The evolution of traditional power grids into smart grids (SGs) has been boosted by advancements in information and communication technologies [1]. SGs are designed to deliver energy that is economical, reliable, secure, and sustainable, while simultaneously optimizing load distribution, monitoring grid performance, and balancing demand and supply. Accurate forecasting of both energy generation and consumption is fundamental for ensuring grid stability and effective scheduling [2]. Power data, a critical component in this process, is inherently a multivariate time series, comprising sequentially ordered data points across multiple interrelated variables. This data reveals two distinct types of patterns: local patterns that capture short-term variations and anomalies, and global patterns that highlight overarching trends in energy generation and consumption. These patterns serve as the foundation for reliable forecasting by encapsulating critical insights into the dynamics of the power grid for electrical automation.

Forecasting energy generation and consumption provides significant advantages for energy providers. By analyzing historical data, providers can predict future trends, enabling the seamless supply of energy to meet demand [3]. This involves not only identifying local and global patterns within time series data but also deciphering the complex inter-dependencies between them. Such insights facilitate proactive decision-making, allowing power suppliers to optimize energy production, minimize costs, and promote environmentally sustainable practices. For instance, a deeper understanding of power generation and consumption patterns supports utility companies in efficiently allocating resources and managing energy systems in smart cities. The interdependence between generation and consumption underscores the importance of unified models for electrical automation, which we define as models capable of jointly predicting both processes to capture their mutual influences. Our proposed dual-stream hybrid model, which integrates Convolutional Neural Networks (CNNs) with spatial attention and Gated Recurrent Units (GRUs) with self-attention, is designed to enhance forecasting accuracy for both power generation and consumption within smart grid automation systems. Accurate power generation forecasts enable grid operators to anticipate energy supply requirements, while precise power consumption predictions assist in managing demand. By integrating predictions for both power generation and power consumption, a comprehensive perspective on electricity grid dynamics emerges, empowering stakeholders to make informed decisions and enhance the overall efficiency of grid operations. This holistic approach ensures that smart grids achieve their goal of delivering sustainable and efficient energy solutions for modern communities [4]. Highly accurate and efficient forecasting methods are essential for effective grid management, enabling electricity companies to make critical decisions [5]. Even a modest improvement, such as reducing forecasting errors by 1%—can result in savings amounting to tens of millions of pounds [6]. The core objective of forecasting power generation and consumption is to ensure a steady power supply while minimizing operational expenses.

The remainder of this paper is organized as follows: Section 2 reviews the existing literature on the topic, Section 3 provides a brief description of the proposed method, Section 4 presentation of the experimental results, where the performance of the proposed method is evaluated in terms of power generation and consumption forecasting, Section 5 discusses the model's performance compared to state-of-the-art methods and its practical implications, and Section 6 concludes the paper.

2 Related work

Despite its importance, accurately predicting power generation and consumption remains a significant challenge. Power generation data is often chaotic, random, and intermittent, making it difficult to predict with precision. Similarly, fluctuations in power consumption arise from the diverse behaviors of consumers, influenced by factors such as special occasions, increased adoption of smart devices, and the growing use of electric vehicles [7]. To address these challenges, accurate forecasting methods are considered as a key strategy for maintaining a reliable power supply. It plays a pivotal role in achieving an equilibrium between power generation and consumption, ultimately ensuring the continuous operation of modern electricity grids. Given the critical role of power generation and consumption forecasting, several models are proposed in the literature, that are categorized into short-, long-, and medium-term forecasting horizons. However, short-term forecasting is crucial for balancing the power. On the other hand, medium- and long-term forecasts are primarily used for settling energy prices, planning load dispatch, and scheduling maintenance behavior. Short-term forecasting finds extensive applications, making it a focal point for researchers aiming to enhance accuracy. The smooth operation of smart grids, particularly through electrical automation, heavily depends on precise short-term forecasts, prompting the development of advanced, data-driven methodologies. These methods are broadly classified into statistical techniques, AI approaches, deep learning models, and hybrid frameworks that combine multiple methodologies to maximize performance. This diversity in approaches reflects ongoing efforts to optimize forecasting techniques and ensure the reliable operation of modern energy systems.

In the early stages of load and demand forecasting, statistical methods were predominantly employed, leading to the development of various models. These include approaches based on Bayesian analysis [8], autoregressive techniques and moving averages [9], Hammerstein frameworks [10], Kalman filters [11], and multiple kernel methods [12]. While these techniques offered valuable insights, their effectiveness diminishes when dealing with nonlinear and complex datasets. To address these limitations, researchers have shifted their focus toward AI-assisted methods. These advanced approaches leverage the strengths of artificial intelligence to overcome the challenges posed by nonlinear data, providing more accurate and robust forecasting solutions. This transition highlights the evolving nature of forecasting techniques as they adapt to the increasing complexity of modern energy systems. These include methods utilizing support vector machines, artificial neural networks, extreme learning machines, and fuzzy networks [13, 14, 15, 16]. While these approaches have demonstrated improved accuracy compared to traditional techniques, they often rely on manual features which limit their ability to generalize across diverse datasets.

A significant limitation of these shallow models lies in their tendency to experience parameter non-convergence and instability when dealing with the vast historical datasets inherent in power data. Recognizing these challenges, researchers and domain experts have increasingly turned to deep learning methodologies. Deep learning offers a robust alternative, leveraging its capacity to learn hierarchical representations directly from raw data, thus addressing the shortcomings of earlier AI-based models and paving the way for more stable and generalizable forecasting solutions. Deep learning-based models have revolutionized forecasting by offering end-to-end feature extraction capabilities and the ability to learn from extensive datasets while maintaining high generalizability. Recent advancements have extended their application to power forecasting, utilizing architectures such as RNNs [31], ESNs [32], LSTM networks [33], GRUs [34], and CNNs [35, 36]. Compared to traditional statistical and shallow learning methods, deep learning models offer significant advantages in handling complex datasets. However, developing effective models requires a deep understanding of the underlying data. In the context of power generation and consumption forecasting, historical data often exhibits time series patterns with spatiotemporal dependencies. While deep learning models excel in learning specific features, their ability to process multiple feature types simultaneously is limited. To address this, hybrid models have emerged, combining the strengths of various architectures to effectively capture the intricate spatiotemporal representations inherent in such datasets.

Hybrid models that integrate multiple techniques have proven highly effective in achieving accurate power generation and consumption forecasting. Notable hybrid approaches include combinations such as CNN-RNN [37], autoencoders with BiLSTM [38], CNN-LSTM [39], CNN-GRU [40], ConvLSTM (CLSTM) [41], and ESN-CNN. Among these, models that pair CNNs with RNN variants, such as LSTM, GRU, and ESN, have consistently demonstrated state-of-the-art performance. These hybrid architectures excel in extracting spatiotemporal features from historical data, enabling precise predictions for power generation and consumption. Recent advancements in hybrid forecasting models have incorporated attention mech-

Table 1: Summary of related work on power forecasting models

Work	Model	Dataset	Metrics	Limitations
Chen et al. [17]	RCC-LSTM (residual CNN + LSTM)	Solar PV generation data (5-min resolution)	RMSE & MAE	Inflexible selection of RCC threshold values
Zang et al. [18]	Hybrid CNN with meta-learning	DKASC	MAE, RMSE reported	Computationally heavy; limited validation; unclear generalizability.
Zhou et al. [19]	ELM + genetic algorithm	PV output (15-min data)	R^2, MAE, nRMSE	Needs more diverse data; may underperform across varied locations.
Cheng et al. [20]	Graph-based DL model	PV plant data (hourly)	RMSE, MAPE	Requires complex preprocessing; tested on narrow scope.
Korkmaz [21]	SolarNet: CNN + VMD	Solar PV (hourly)	RMSE=0.309, MAE=0.175	Single-site test; preprocessing adds overhead; uncertain performance in extremes.
Wang, Qi, and Liu [22]	Deep Neural Networks	1B DKASC	RMSE, MAPE (improvements)	risks overfitting; lacks uncertainty modeling.
Rajabi and Estebsari [23]	2D-CNN on recurrence plots	IHEPC	RMSE, MAE (20% RMSE vs 1D CNN)	Small-scale test; Limited exploration of time series to image encoding methods.
Ullah et al. [24]	CNN + Bi-LSTM	Residential load dataset	RMSE, MAE	Short-term focus; smart meter data required; only deployable on PCs.
Haq et al. [25]	ConvLSTM-BiLSTM	Res. & comm. load (smart grid)	RMSE, MAPE (improvements)	complex and resource-heavy; Not designed for deployment on resource-constrained devices.
Kim and Cho [26]	CNN-LSTM hybrid	IHEPC	RMSE, MAE	Limited Dataset diversity
Abdel-Basset et al. [27]	STLF-Net	IHEPC and AEP	RMSE, MAE, MAPE	Data ambiguities/irregularities may limit real-world accuracy/confidence. Load data vulnerability to attacks.
Han et al. [28]	Multilayer GRU	IoT energy data	MSE, RMSE	Primarily focused on short-term load forecasting (STLF). edge device constraints.
Kim and Cho [29]	Autoencoder based	Household power consumption	MSE, MRE	Limited data;
Khan et al. [30]	CNN with LSTM-AE	UCI Residential & Korean Commercial building data	MAE, MSE, MAPE	complex model; tested on limited geography.

anisms to further improve accuracy. Examples include transformer-based networks, multi-attention networks, and temporal CNNs with channel-attention [42, 43, 44, 45]. These cutting-edge methods effectively capture complex relationships within the data, leading to significant improvements in forecasting performance. To provide a comprehensive overview of the state-of-the-art, Table 1 summarizes key prior works on power forecasting, comparing their model types, datasets, evaluation metrics, and limitations. The table highlights challenges such as computational complexity, limited dataset diversity, and lack of attention mechanisms in existing models, which often hinder their ability to effectively capture both spatial and temporal dependencies.

This paper emphasizes the following key areas that need further investigation to enhance forecasting accuracy for power generation and consumption:

1. Previous research involving stacked methods has shown better performance; however, these models primarily use previous data that represents power generation or consumption over one or two hours as input. These models typically generate forecasts for the following hour using the sliding window technique. Unlike video and image analysis, power data often lacks the rich feature dimensions necessary for capturing both spatial-temporal relationships efficiently within this architecture.

2. Many existing models process features through their architectures without adequately refining the extracted spatial and temporal representations, often routing unoptimized feature vectors directly to dense layers. While some studies have incorporated attention mechanisms to improve results, these are typically limited

to either spatial or temporal attention, failing to address the combined spatiotemporal nature of power data. Thus, a comprehensive attention mechanism that refines both spatial and temporal features across the network is essential for more accurate forecasting.

3. Another main challenge is the evaluation of the models on specific types of datasets. The lack of varied and high-resolution datasets in evaluations limits the generalization and robustness of these forecasting models.

Addressing these gaps is crucial for advancing the field of power generation and consumption forecasting and improving the accuracy, applicability, and scalability of predictive models. To guide this study, we pose the following research questions:

1. Can our dual-stream hybrid model, as a unified approach for power generation and consumption forecasting, improve accuracy over single-stream deep learning models?

2. How does incorporating spatial and self-attention mechanisms improve forecasting performance?

These questions evaluate the effectiveness of the proposed dual-stream hybrid model, which processes both generation and consumption data to capture their inter-dependencies, and the role of attention mechanisms in enhancing forecasting accuracy. This paper presents several key contributions:

1. A dual-stream architecture is developed for power forecasting, processing multivariate time series data in parallel through CNN and GRU streams. This approach efficiently captures spatiotemporal dependencies, addressing the limitations of stacked models that struggle with limited feature dimensions in power data.

2. To enhance feature refinement across the network, we integrate a Spatial Attention Module (SAM) in the CNN stream to prioritize critical spatial features and a Self-Attention Module (SEAM) in the GRU stream to refine temporal dependencies. These modules dynamically select and emphasize relevant features, overcoming the shortcomings of models that lack comprehensive feature refinement.

3. The proposed model undergoes rigorous evaluation across multiple datasets. The results consistently demonstrate that our model outperforms previous baseline approaches in both power generation and consumption forecasting, showcasing its versatility and reliability across different contexts.

3 Proposed solution

The choice of CNNs for spatial feature extraction and GRUs for temporal dependency modeling in the proposed dual-stream architecture is motivated by their complementary strengths in handling the complex spatiotemporal characteristics of power data. CNNs are highly effective in capturing spatial patterns due to their ability to apply convolutional filters that identify local correlations and hierarchical features within multivariate time series data, such as weather-related variables (e.g., temperature, humidity) influencing power generation. This makes CNNs particularly suitable for extracting spatial dependencies in power data, which often exhibit localized patterns driven by environmental factors. In contrast, GRUs are designed to model sequential data by maintaining memory of previous time steps through update and reset gates. Compared to LSTM networks, GRUs offer a simpler architecture with fewer parameters, making them computationally efficient while still effectively capturing long-term temporal dependencies in time series data. The decision to use GRUs over other RNN variants is driven by their balance of performance and computational efficiency, which is critical for real-time forecasting applications in electrical automation systems. This dual-stream approach leverages the strengths of both architectures to address the limitations of single-stream models, which often struggle to simultaneously capture spatial and temporal features.

Likewise, the selection of a dual-stream CNN-GRU architecture over other alternatives such as transformer-based models or hybrid BiLSTM-attention models is driven by several key considerations. Transformer-based models, while powerful for capturing long-range dependencies, require significant computational resources and large datasets for effective training, which may not be ideal for real-time forecasting in resource-constrained environments like microgrids. Additionally, transformers primarily focus on temporal relationships and may not efficiently capture spatial dependencies in multivariate power data without complex modifications. The dual-stream approach allows parallel processing of spatial and temporal features, enhancing the model's ability to capture the interdependence between power generation and consumption. This design choice aligns with the need for a unified model that jointly predicts both processes while maintaining computational efficiency.

The core framework of the proposed Dual-Stream Network is illustrated in Figure 1. In this architecture, the input data is processed in parallel through the two streams. These architectures are designed to extract spatiotemporal features from the input data. The outputs of both the CNN and GRU streams are then merged into a single feature vector, which is used for further processing. To optimize the feature selection, a SAM is incorporated to identify the most representative features, which are then utilized for the final forecasting task. The following subsections provide an in-depth description of the internal components of the proposed model.

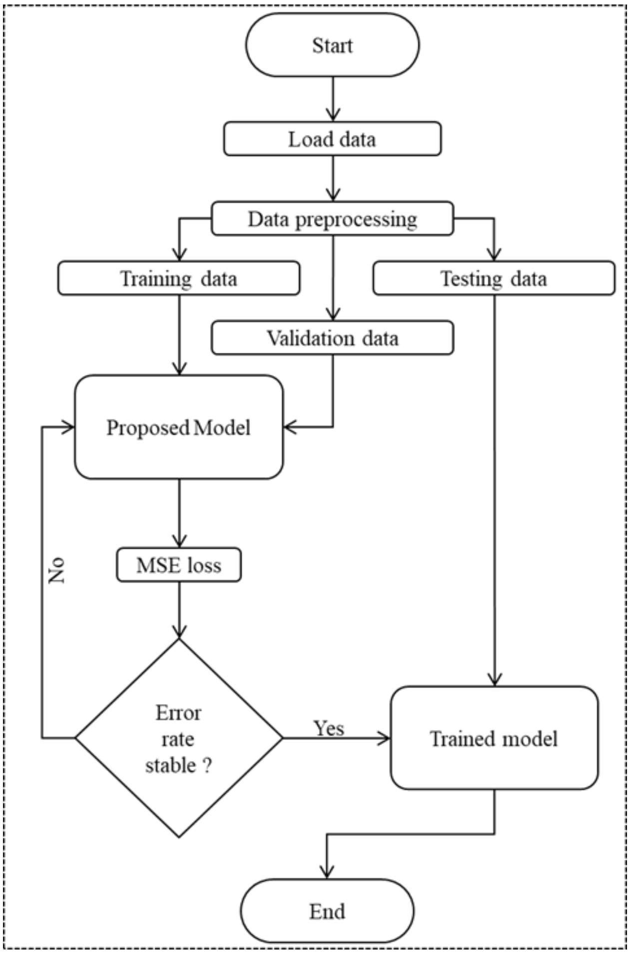

Figure 1: Flow diagram of proposed model

3.1 Dual stream network

The proposed Dual-Stream Network (DSN) architecture consists of two parallel streams, one using CNN layers for extracting cross-variable correlations and the other employing GRU layers for capturing temporal dependencies. Together, these streams are designed to handle the complex spatiotemporal characteristics of power data, which is manipulated by different weather-related factors such as temperature, humidity, and rain. The CNN stream processes a multivariate input sequence of PV power and weather-related variables, structured as a 2D matrix (time steps × variables, e.g., PV power, temperature, humidity). In this context, the variables represent a pseudo-spatial dimension, where the CNN captures correlations among them, such as the influence of temperature and humidity on PV power output. These cross-variable relationships are analogous to spatial patterns in image processing, enabling the CNN to extract meaningful features for power forecasting. The input to the CNN consists of a feature vector $X_i^0 = X_1, X_2, X_3 \ldots . X_n$ where each element corresponds to a variable, and the data is passed through different CNN layers and pooling layers. The CNN applies convolutional filters to produce feature maps that encode these cross-variable correlations, which are subsequently refined by SAM to prioritize the most relevant features (e.g., emphasizing temperature over less impactful variables). The convolution operation is used to extract cross-variable patterns in the data by applying filters to the input sequence. The output feature vector from the convolution layer $Y_{i,j}^1$ is calculated as:

$$Y_{i,j}^1 = \alpha(B_j^1 + \sum_{m=1}^{M} W_{m,j}^1 X_{i+m-1,j}^0 \qquad (1)$$

Here, B_j^1 represents the bias term, $W_{m,j}^1$ is the weight matrix, and α is the activation function. The convolutional layers are followed by max pooling to reduce the dimensionality of the output and retain the most important features:

$$Y_{i,j}^l = \alpha(B_j^l + \sum_{m=1}^{M} W_{m,j}^l X_{i+m-1,j}^0 \qquad (2)$$

The GRU stream is designed to handle temporal dependencies by using GRU layers, which are lighter and more efficient compared to traditional RNNs. The GRU processes the same multivariate time series input, but focuses on the temporal dimension, modeling sequential patterns across time steps (e.g., daily or seasonal trends in power consumption). The GRU outputs a sequence of hidden states capturing these temporal dependencies, which are refined by SEAM to focus on significant time steps. The GRU architecture consists of two main gates: the Reset Gate and the Update Gate. The update gate u_t is calculated as:

$$u_t = \alpha[(w_1 \times i_t) + (w_2 \times i_{t-1})] \qquad (3)$$

The reset-gate r_t find out the degree to which the previous hidden state is remembered, and is calculated as:

$$r_t = \alpha[(w_1 \times i_t) + (w_2 \times i_{t-1})] \qquad (4)$$

The memory content $m_r^{`}$ captures the past information and is computed as:

$$m_r^{`} = \tanh[(w_1 \times i_t) + (r_t \cdot i_{t-1})] \qquad (5)$$

Finally, the output memory combines previous hidden state and the memory content:

$$m_t = (u_t \cdot i_{t-1} + (1 - u_t) \cdot m_r^{`}) \qquad (6)$$

Spatial Attention: The extracted feature vector is passed to the 1D-SAM, which refines feature representation by emphasizing the most important elements. SAM operates on the CNN output, which consists of feature maps representing spatial correlations among input variables (e.g., weather-related features like temperature and humidity) across time steps. It creates a spatial attention map by analyzing these inter-variable relationships, prioritizing features that are most relevant to power forecasting. This module creates a spatial attention map by analyzing the inter-spatial relationships within the feature data. To achieve this, it applies avg-pooling and max-pooling along the channel-dimension, producing two distinct feature maps: $\mathrm{M}_s(\mathrm{F}) \in \mathrm{R}^{t \times c}$, where t represents the time steps and c indicates the number of channels. These pooled feature maps are then concatenated and processed through a 1D convolutional layer with a kernel size 3×3, resulting in the final spatial attention map $\mathrm{M}_s(\mathrm{F}) \in \mathrm{R}^{t \times c}$. This attention map assigns importance to specific regions in the feature space, effectively determining which features should be enhanced or suppressed as given below:

$$\mathrm{M}_s(\mathrm{F}) = \sigma(f^{3\times3}([\mathrm{AvgP}(\mathrm{l}_5), \mathrm{MaxP}(\mathrm{l}_5)])) \tag{7}$$

$$\mathrm{M}_s(\mathrm{F}) = \sigma(f^{3\times3}(\mathrm{l}^s_{5\mathrm{avg}}, \mathrm{Fl}^s_{5\mathrm{max}})) \tag{8}$$

Here, σ denotes the sigmoid activation function, $f^{(3\times3)}$ is the convolutional operation with a kernel size of 3×3), and AvgP and MaxP refer to the average-pooling and max-pooling operations, respectively. This process ensures that SAM enhances the model's focus on critical spatial features, such as weather variables strongly correlated with power generation, while suppressing less relevant ones, thereby improving forecasting accuracy.

Self-Attention: The SEAM processes the output of the GRU stream, which consists of a sequence of hidden states capturing temporal dependencies across time steps. Unlike SAM, which focuses on spatial correlations among variables, SEAM dynamically selects the most relevant temporal features by assigning weights to significant time steps in the sequence. The attention scores for each feature in the J^{th} timestamp in the d^{th} dimension is calculated as:

$$S_{J,d} = f_{\mathrm{sco}}(W_{J,d}[h_{1,d}, h_{2,d}, h_{3,d}, \ldots, h_{n,d}]) \tag{9}$$

where $d = 1, 2, 3, \ldots, n$ and $J = 1, 2, 3, \ldots, o$. Here, $h_{J,d}$ represents the hidden state at the J^{th} timestamp and the d^{th} dimension, while f_{sco} is a function (implemented with fully connected layers) to score the importance of each feature. This mechanism allows SEAM to emphasize critical temporal patterns, such as peak consumption periods, while down-weighting less significant time steps, complementing SAM's spatial refinement.

Finally, the output from the SAM and SEAM is merged and forwarded to Dense layers, where the final power forecast is generated. The fully connected layers use the following operation to calculate the output d_i^l:

$$d_i^l = \sum_j W_{ji}^{l-1}(\alpha(X_i^{l-1}) + B_j^{l-1}) \tag{10}$$

Where W_{ji}^{l-1} is the weight matrix, α is the activation function, and B_j^{l-1} is the bias term. This architecture ensures effective extraction of both spatial and temporal features while utilizing attention mechanisms to improve feature selection for more accurate forecasting. The model is capable of handling complex, irregular patterns in PV power generation, resulting in a robust and accurate forecasting system.

3.2 Network architecture

The network of the model is proposed to efficiently capture spatiotemporal features for precise forecasting of power. The network consists of convolutional, GRU, spatial attention, Self-attention, and fully connected layers. The model's parameters are adjusted through extensive experimentation to achieve optimal performance. This customization ensures that the architecture is fine-tuned for accurate forecasting. The input to the model is 12×10 (for power generation) data matrix, which contains the necessary weather-related variables (e.g., temperature, humidity, rain, etc.) and power generation or consumption. This data is passed through both streams in parallel to learn spatial-temporal features. The first stream utilized two convolutional layers and a pooling layer. This setup allows for spatial feature extraction from the input data with Filter sizes: 16 and 32, 3 Kernel size, and ReLU activation-function. The max pooling operation is used to reduce dimensionality and retain important features. The output of CNN is then passed through SAM to select optimal features in spatial dimension. The GRU stream consists of two stacked GRU layers to capture temporal dependencies from the input data with Cell sizes: 32 and 16 (for each GRU layer). GRU layers are particularly effective at learning sequential patterns and trends in time series data. The output of GRU is then passed through the SEAM, which dynamically selects the most important features from the fused vector by assigning different weights to different features. The SEAM helps highlight critical temporal and spatial features while down-weighting less important ones. Finally, the output of spatial attention send self-attention is fused and passed through fully connected layers to generate the final power forecast. Extensive experimentation is conducted to finalize the model's internal parameters. The experiments ensure that the model's performance is optimized for accurate and reliable forecasting. The parameters, including the filter sizes, kernel sizes, GRU cell sizes, and other network settings, are selected based on their ability to extract meaningful features and make accurate predictions. A flow diagram of the proposed model is illustrated in Figure 1.

3.2.1 Hyperparameter selection and tuning

To ensure optimal performance of the proposed dual-stream CNN-GRU model, a systematic hyperparameter selection and tuning process was employed. The CNN stream consists of two convolutional layers with 16 and 32 filters, respectively, each using a kernel size of 3 and ReLU activation, followed by a max-pooling layer with a pool size of 2. The GRU stream comprises two stacked GRU layers with 32 and 16 cells, respectively, using tanh activation for the recurrent steps. The fully connected layers include a hidden layer with 64 neurons (ReLU activation) and an output layer with 1 neuron (linear activation). The Spatial Attention Module (SAM) uses a 1D convolutional layer with a 3×3 kernel, and the Self-Attention Module (SEAM) employs a fully connected layer to compute attention scores.

Hyperparameter tuning was conducted using a grid search over the following ranges: CNN filter sizes ([8, 16], [16, 32], [32, 64]), GRU cell sizes ([16, 8], [32, 16], [64, 32]), learning rates ([0.0001, 0.001, 0.01]), and batch sizes ([32, 64, 128]). The Adam optimization algorithm was selected for its robustness in training deep neural networks, with a default beta1 of 0.9 and beta2 of 0.999. The model was trained for 100 epochs with early stopping (patience of 10 epochs) to prevent overfitting. The tuning process utilized 5-fold cross-validation on the DKASC dataset to evaluate performance, with the root mean squared error (RMSE) on the validation set as the primary metric. This configuration was further validated on the IHEPC dataset to ensure generalizability. The systematic tuning process ensures transparency and reproducibility, providing a robust foundation for the model's performance in power forecasting applications.

4 Results

The experimental results are presented for power generation and consumption prediction. We utilized datasets containing time series data on factors such as weather and time of day for solar power generation and electricity consumption. An ablation study was also conducted to examine the contributions of different components of the model. The results are further analyzed and compared with state-of-the-art methods in the Discussion section.

4.1 Datasets

The electricity generation and consumption datasets used in this study include data for both renewable power generation and household electricity consumption. For renewable power generation, the DKASC datasets are utilized, which includes daily data recorded at a five-minute resolution from multiple active solar power plants in Alice Springs, Australia. The three datasets used are Trina-23.4-kW, Trina-10.5-kW, and Eco-Kinetics-26.5-kW. These datasets contain both renewable power generation and weather-related variables like humidity, rainfall, and temperature. For electricity consumption prediction, the IHEPC dataset is used, which includes one-minute resolution data from 2006 to 2010. This dataset records information on active and reactive power, voltage, intensity, and submetering, along with time and date information.

While the DKASC and IHEPC datasets provide valuable insights into renewable power generation and household electricity consumption, they have certain limitations that impact their representativeness of real-world power grids. The DKASC datasets, collected from solar power plants in Alice Springs, Australia, primarily reflect the climatic and operational conditions of a single geographical region with a arid climate, which may not fully capture the variability of solar generation in regions with different weather patterns, such as tropical or temperate climates. Similarly, the IHEPC dataset, derived from a single household in France, may not adequately represent the diverse consumption patterns observed in larger, industrial, or multi-household grid configurations. These datasets also lack significant variability in grid infrastructure, such as differences in energy storage systems or hybrid renewable sources, which are common in modern smart grids. Despite these constraints, the datasets' high temporal resolution (5-minute for DKASC, 1-minute for IHEPC) and inclusion of relevant variables (e.g., weather, voltage) make them suitable for developing and testing forecasting models, though broader applicability requires further validation across diverse datasets.

4.2 Data preprocessing

To ensure the quality and consistency of the DKASC and IHEPC datasets for model training and evaluation, several preprocessing steps were applied. First, data normalization was performed using min-max scaling to transform all features (e.g., power generation, consumption, temperature, humidity) into a [0, 1] range. This step mitigates the impact of varying scales across different variables, ensuring stable training of the neural network. The normalization formula used is given by:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{11}$$

where x is the original value, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the feature across the dataset. Missing data, which accounted for less than 2% of the records in both datasets, were handled using linear interpolation for time-series continuity, leveraging neighboring data points to estimate missing values. This approach was chosen to preserve temporal patterns critical for forecasting. Outliers, identified as values exceeding three standard deviations from the mean, were capped at the 99th percentile to reduce their influence without removing data points. The datasets were partitioned into training, validation, and test sets using a 70:15:15 split, respectively. To maintain temporal order, the data was split sequentially, with earlier data used for training and later data for validation and testing. This partitioning ensures that the model is evaluated on unseen data, simulating real-world forecasting scenarios.

4.3 Evaluation metrics

The performance of the proposed model is evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (nRMSE), and Mean Absolute Error (MAE), as these metrics are widely adopted in time-series forecasting, particularly for power generation and consumption. MSE and RMSE are selected because they emphasize larger errors, which is critical in power forecasting where significant deviations can lead to costly grid imbalances. RMSE provides an interpretable measure in the same units as the forecasted variable, making it suitable for assessing the magnitude of pre-

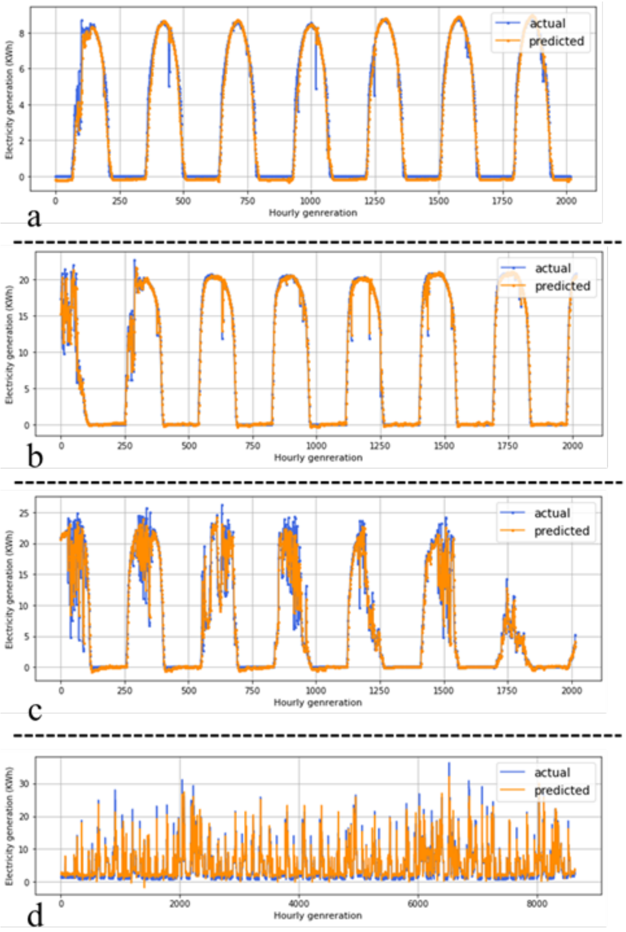

Figure 2: Predicted vs actual values where a, b, c show power generation and d show power consumption results

diction errors. However, both MSE and RMSE are sensitive to outliers, which can skew results in datasets with extreme values, such as those caused by sudden weather changes. nRMSE normalizes RMSE by the range of the target variable, enabling fair comparisons across datasets with different scales, which is essential for evaluating model performance across diverse power generation and consumption datasets. MAE, on the other hand, measures the average absolute error, offering robustness to outliers and a straightforward interpretation of average prediction accuracy. While MAE is less sensitive to large errors compared to RMSE, it complements the other metrics by providing a balanced perspective on model performance. Together, these metrics provide a comprehensive evaluation of the model's accuracy and robustness, aligning with the requirements of power forecasting applications where both small and large errors impact grid reliability. The equations for these metrics are given below:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(\alpha_i - p_i)^2}{N}} \tag{12}$$

$$\text{MSE} = \frac{\sum_{i=1}^{m}(\alpha_i - p_i)^2}{N} \tag{13}$$

$$\text{nRMSE} = \text{RMSE}/(\alpha_{\text{max}} - \alpha_{\text{min}}) \tag{14}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|\alpha_i - p_i|}{N} \tag{15}$$

where α and p show the actual and model prediction.

4.4 Ablation study

The performance of the proposed model is evaluated by comparing it with several predictive modeling techniques, including traditional regression methods (SVR, Decision Tree, Linear Regression), deep models (MLP, LSTM, CNN), a transformer-based model, and a hybrid BiLSTM-attention model during the ablation study. The transformer-based model employs a multi-head self-attention mechanism to capture temporal dependencies, while the BiLSTM-attention model combines bidirectional LSTM layers with an attention mechanism to focus on relevant temporal features. Experimental results presented in Figures 3 to 6 demonstrate that traditional regression models perform worse than deep learning-based models. In particular, hybrid ones perform better comparatively. Among the hybrid models, the proposed model achieved the lowest error rates. For instance, on the Tarina 10.5 kW dataset, the proposed model achieved an RMSE of 0.0985, compared to 0.124 for the transformer-based model and 0.109 for the BiLSTM-attention model, highlighting the dual-stream architecture's superior ability to capture spatiotemporal dependencies. Similar trends were observed across other datasets, with the transformer model exhibiting higher computational costs (e.g., 2.15 seconds inference time on GPU) and the BiLSTM-attention model showing slightly higher errors due to its focus on temporal rather than spatial features. For example, on the Tarina 10.5 kW dataset, the proposed model achieved RMSE, MSE, nRMSE, and MAE values of 0.0985, 0.0097, 0.1198, and 0.0490, respectively. For the Tarina 23.4 kW dataset, the values were 0.0574 RMSE, 0.0033 MSE, 0.1512 nRMSE, and 0.0430 MAE, while for the Eco-Kinetics 26.5 kW dataset, the model achieved 0.0346 RMSE, 0.0012 MSE, 0.2393 nRMSE, and 0.0128 (MAE). The proposed model also outperformed others on the IHEPC data, with RMSE, MSE, nRMSE, and MAE values of 0.0300, 0.0009, 0.1158, and 0.0166, respectively. All experiments were conducted for one-hour-ahead predictions, and the proposed model consistently showed the lowest error rates, as evidenced by the close alignment between predicted and actual values in Figure 2. This confirms the model's applicability for both renewable power generation and electricity consumption prediction.

5 Discussion

The performance of the proposed dual-stream CNN-GRU model is analyzed compared to state-of-the-art (SOTA) methods, exploring the reasons behind its superior performance, and the practical implications of the results for electrical automation and smart grid applications. Additionally,

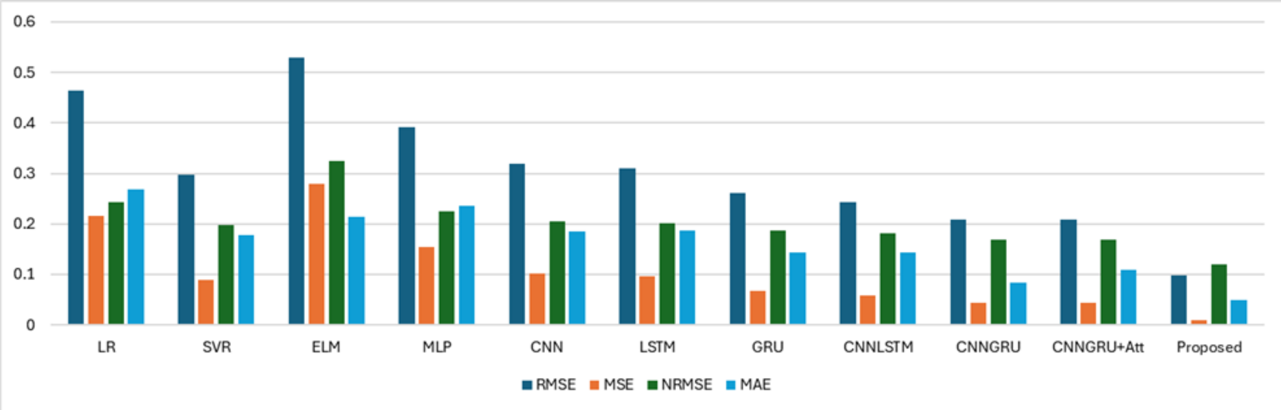

Figure 3: Comparative analysis of different models developed during ablation study over Tarina 10.5 kW

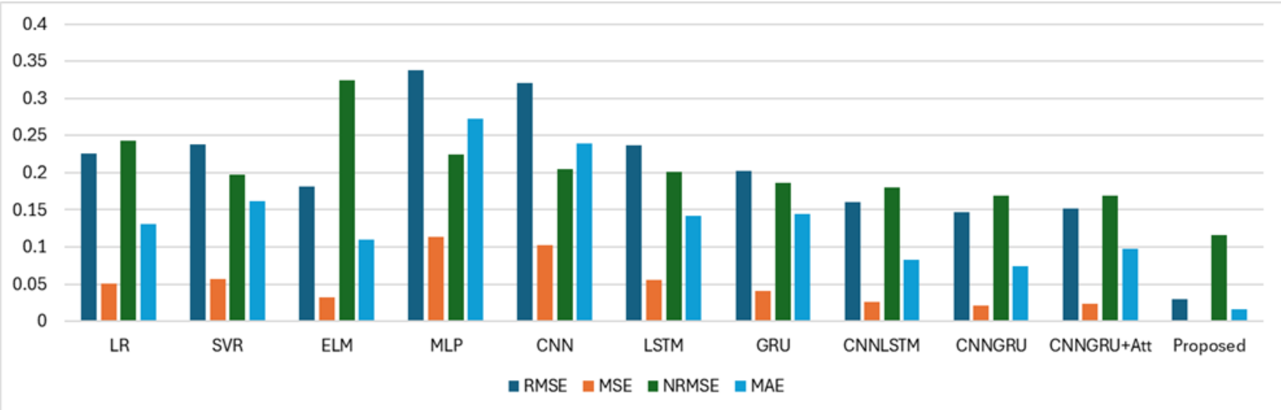

Figure 4: Comparative analysis of different models developed during ablation study over IHEPC

we discuss its scalability and deployment feasibility to address real-world constraints.

5.1 Comparative analysis of the proposed model with state-of-the-art models

The effectiveness of the proposed model is evaluated on the mentioned datasets, and its performance is compared with several baseline methods. For DKASC, the proposed model demonstrates significant improvements over the baseline methods, as summarized in Table 2. The RMSE, MSE, and MAE values achieved by the proposed model are 0.0635, 0.0047, and 0.0349, respectively. These results indicate a substantial reduction in prediction error compared to other state-of-the-art models. For instance, Chen et al. [17] reported an RMSE of 0.94 and an MAE of 0.587, while Zang et al. [18] achieved an MSE of 0.081 and an MAE of 0.152. The proposed model's superior performance can be attributed to its dual-stream architecture unlike single-stream models like Chen et al. [17] that rely on stacked layers. The integration of spatial and self-attention mechanisms further refines feature selection, enabling the model to prioritize relevant patterns, unlike Zang et al. [18], which lacks such mechanisms and suffers from computational complexity. However, on datasets with high variability (e.g., Eco-Kinetics 26.5 kW, nRMSE=0.2393), the model's performance, while still superior, shows a slightly higher normalized error, possibly due to the influence of extreme weather conditions that challenge even advanced feature extraction.

To validate the statistical significance of these improvements, we conducted Wilcoxon signed-rank tests across multiple runs on DKASC and IHEPC datasets. The tests yielded p-values < 0.05, indicating that the proposed model's improvements are statistically significant. Additionally, 95% confidence intervals for the proposed model's RMSE on the DKASC dataset (0.059–0.068) and IHEPC dataset (0.027–0.033) confirm the reliability of the reported metrics, with narrow intervals reflecting consistent performance across runs. These statistical analyses strengthen the claim that the dual-stream architecture and attention mechanisms contribute to robust forecasting improvements over SOTA methods.

On the IHEPC dataset, the proposed model also outperforms existing techniques, achieving an RMSE of 0.0300, an MSE of 0.0009, and an MAE of 0.0166, as shown in Table 3. These metrics are significantly better than those of the competing methods, such as Rajabi and Estebsari

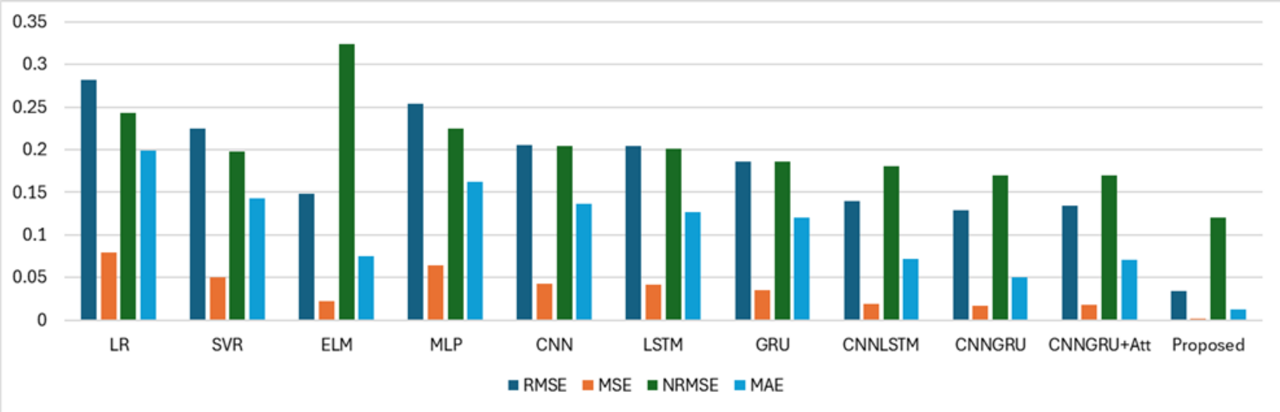

Figure 5: Comparative analysis of different models developed during ablation study over Tarina 26.5 kW

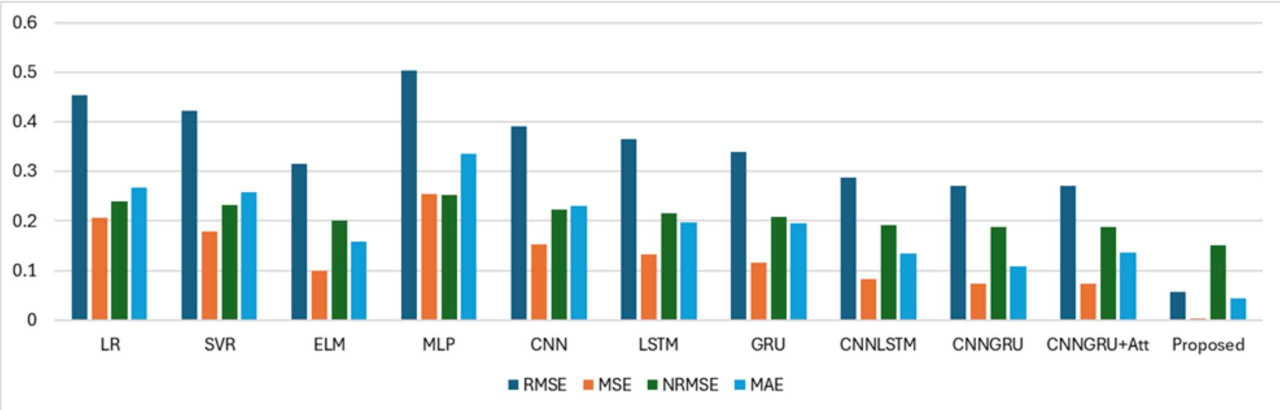

Figure 6: Comparative analysis of different models developed during ablation study over Tarina 23.4 kW

[23], who reported an RMSE of 0.79 and an MAE of 0.59, and Haq et al. [25], with an RMSE of 0.32 and an MAE of 0.31. The model's success on the IHEPC dataset is driven by the GRU stream's ability to capture long-term temporal dependencies in high-resolution (1-minute) consumption data, complemented by the CNN stream's extraction of spatial patterns from correlated variables like voltage and intensity. In contrast, models like Rajabi and Estebsari [23] rely on 2D-CNNs with limited temporal modeling, leading to higher errors. The attention mechanisms mitigate the impact of data irregularities noted in Abdel-Basset et al. [27], enhancing robustness. However, the model's performance advantage is less pronounced on datasets with significant noise, suggesting a potential area for improvement in handling data ambiguities.

5.2 Time complexity analysis

To ensure the proposed model is suitable for real-time power forecasting, we evaluate its computational efficiency by measuring inference time, defined as the time required for a single forward pass to generate a prediction. Inference time is critical for real-time applications in electrical automation, where rapid predictions are necessary to maintain grid stability. We compare the inference time of the proposed model against baseline models on three platforms-CPU, GPU, and Raspberry Pi (RPI)—to demonstrate its efficiency across diverse computational environments, including edge devices like Raspberry Pi commonly used in microgrids. All times are reported in seconds, ensuring consistency and clarity. This comparison highlights the model's lightweight design, which minimizes computational costs and response delays, thereby enhancing its applicability in real-time energy management systems.

To address this, we conduct a time complexity analysis of the model using 3 settings: CPU, GPU, and RPI as given in Table 4, whereas the results are given in Table 5. From the analysis, it is evident that our model has lower inference time than others across all platforms. Specifically, the proposed model achieves an inference time of 0.4354 seconds on GPU, 0.9874 seconds on CPU, and 1.674 seconds on RPI. This efficiency stems from the streamlined dual-stream architecture and the use of GRUs, which are computationally lighter than LSTMs used in models like Haq et al. [25]. In contrast, models like Han et al. [28] exhibit significantly higher inference times (e.g., 20.36 seconds on RPI), limiting their suitability for edge devices.

Comparing this with the performance of other models, such as those from Chen et al. [17] and Wang, Qi, and

Table 2: Comparative analysis of the model over the solar dataset

Method	MSE	RMSE	MAE
Chen et al. [17]	-	0.94	0.587
Zang et al. [18]	0.081	-	0.152
Li et al. [46]	-	-	0.2805
Zhou et al. [19]	-	-	0.2367
Cheng et al. [20]	-	0.336	0.177
Li et al. [47]	-	0.2357	-
Korkmaz [21]	-	0.309	0.175
Wang, Qi, and Liu [22]	-	0.343	0.126
Wang, Qi, and Liu [48]	-	0.621	0.221
Ours	**0.0047**	**0.0635**	**0.0349**

Table 3: Comparative analysis of the model over the IHEPC dataset

Paper	MAE	RMSE	MSE
Rajabi and Estebsari [23]	0.59	0.79	-
Ullah et al. [24]	0.3469	0.5650	0.3193
Haq et al. [25]	0.31	0.32	0.10
Kim and Cho [26]	0.3317	0.5957	0.3549
Khan et al. [30]	0.31	0.47	0.19
Han et al. [28]	0.19	0.22	0.17
Abdel-Basset et al. [27]	0.2674	0.4386	0.1924
Kim and Cho [29]	0.3953	-	0.3840
Khan et al. [49]	0.29	0.42	0.18
Khan et al. [50]	0.0038	0.0614	0.0537
Ours	**0.0009**	**0.0300**	**0.0166**

Table 4: Hardware configuration

Setting	Model	Memory
GPU	GeForce-RTX-3090	24GB
RPI	RPI 4 B+	8GB
CPU	Intel Core i5-6600	32GB

Liu [48], the proposed model's lower inference times (e.g., 0.4354 seconds on GPU vs. 0.72 seconds for Haq et al. [25]) highlight its suitability for real-time applications. The model's efficiency makes it viable for deployment on resource-constrained devices, unlike Haq et al. [25], which is resource-heavy.

In the context of real-world smart grid environments, the proposed model's inference time of 1.674 seconds on RPI is well within the requirements for one-hour-ahead forecasting, where predictions are typically needed every few minutes to support grid operations. This efficiency enables real-time deployment in decentralized microgrids, allowing rapid adjustments to energy production and consumption, thereby enhancing grid reliability and reducing operational delays. However, for applications requiring sub-second predictions, further optimization of the model's architecture may be necessary.

Method	Remarks	CPU	GPU	RPI
Chen et al. [17]	Intel-Core-i5, 8GB-RAM	6.387	-	-
Cheng et al. [20]	GTX-1080-GPU	-	3.5	-
Wang, Qi, and Liu [22]	Intel-Core-i5 8-GB-RAM	0.6217	-	-
Wang, Qi, and Liu [48]	Intel-Core-i5 8-GB-RAM	7.196	-	-
Haq et al. [25]	GeForce-RTX-2070	1.44	0.72	-
Han et al. [28]	Intel-Core-i9 RPI- Cortex-A53	6.38	-	20.36
Ours	**Table 3**	**0.9874**	**0.4354**	**1.674**

Table 5: Inference time (seconds) comparison with baseline models

5.3 Scalability and deployment considerations

The scalability and deployment feasibility of the proposed dual-stream CNN-GRU model are critical for its adoption in large-scale electrical automation systems and resource-constrained environments, such as embedded systems in microgrids. The model's memory footprint is approximately 120 MB, calculated based on the parameters of the CNN (16 and 32 filters), GRU (32 and 16 cells), and fully connected layers. This compact size makes it suitable for deployment on edge devices like the Raspberry Pi 4 B+ (8GB RAM), as demonstrated by its inference time of 1.674 seconds on this platform (Table 5). In large-scale systems, where thousands of nodes may require simultaneous forecasting, the model's parallel dual-stream architecture allows efficient batch processing, reducing computational overhead compared to sequential models like Chen et al. [17].

However, there are trade-offs between accuracy and computational efficiency. The proposed model achieves high accuracy (e.g., RMSE=0.0635 on DKASC) by leveraging spatial and self-attention mechanisms, which increase computational complexity compared to simpler models like linear regression or single-layer MLPs. In resource-constrained environments, such as embedded systems with limited memory (e.g., <2GB RAM), this complexity may necessitate model pruning or quantization to reduce the memory footprint and inference time further. For instance, reducing the number of filters in the CNN stream or GRU cells could lower the memory requirement to below 80 MB, at the cost of a potential 5–10% increase in RMSE, based on preliminary ablation studies. Alternatively, deploying the model on cloud-based infrastructure with GPUs (e.g., GeForce RTX-3090, Table 4) ensures high accuracy and low inference time (0.4354 seconds) for large-scale systems, but this increases operational costs and latency in remote areas with limited connectivity.

To enhance scalability, the model supports distributed deployment, where multiple instances can run on edge devices for local forecasting, with periodic synchronization to a central server for global grid optimization. This approach mitigates bottlenecks in large-scale smart grids. Future work could explore techniques like knowledge distillation to create lighter models for ultra-constrained devices or federated learning to improve scalability across distributed grid networks, ensuring both accuracy and efficiency in diverse deployment scenarios.

5.4 Practical implications

The proposed model's high accuracy (e.g., RMSE=0.0635 on the DKASC dataset) and low inference time (e.g., 1.674 seconds on Raspberry Pi) have significant implications for electrical automation and smart grid management. By providing precise one-hour-ahead forecasts, the model enables grid operators to optimize energy production and distribution, reducing operational costs and enhancing grid stability. For instance, accurate solar power predictions allow for better integration of renewable energy, minimizing reliance on fossil fuel backups. The model's compatibility with edge devices like RPI supports decentralized energy management in microgrids, enabling real-time decision-making in smart cities. However, challenges such as handling extreme weather-induced data variability suggest the need for future enhancements, such as incorporating uncertainty modeling to further improve robustness.

6 Conclusion

The primary goal of the smart grid is to maintain equilibrium between power, enabling effective and dependable energy management. Predictive modeling plays a crucial role in achieving this balance, providing insights into future electricity generation and consumption patterns. While numerous predictive models have been developed in the literature, there remains a need for further improvement in prediction accuracy to enhance the energy sector with lower computation. This study introduces a robust and efficient hybrid model designed to forecast electricity generation and consumption effectively. The proposed model integrates CNN with spatial attention and GRU with self-attention in dual stream mechanism, achieving high prediction accuracy while maintaining low computational complexity. After conducting a comprehensive study of various models, the results confirm that the proposed model outperforms baselines in terms of performance and execution time. In the future, the focus will shift towards investigating cutting-edge technologies like reinforcement learning, explainable AI, active learning, and lifelong learning methods to enhance the accuracy and effectiveness of forecasting models. These innovations will play a crucial role in creating more efficient and flexible systems, enhancing the adaptability of smart grid applications for future energy management needs.

References

[1] Jie Ma and Xiandong Ma. "A review of forecasting algorithms and energy management strategies for microgrids". In: *Systems Science; Control Engineering* 6.1 (Jan. 2018), pp. 237–248. ISSN: 2164-2583. DOI: 10.1080/21642583.2018.1480979. URL: http://dx.doi.org/10.1080/21642583.2018.1480979.

[2] Yu Wang et al. "LASSO and LSTM Integrated Temporal Model for Short-Term Solar Intensity Forecasting". In: *IEEE Internet of Things Journal* 6.2 (Apr. 2019), pp. 2933–2944. ISSN: 2372-2541. DOI: 10.1109/jiot.2018.2877510. URL: http://dx.doi.org/10.1109/jiot.2018.2877510.

[3] Guobin Xu et al. "Towards Integrating Distributed Energy Resources and Storage Devices in Smart Grid". In: *IEEE Internet of Things Journal* (2016), pp. 1–1. ISSN: 2327-4662. DOI: 10.1109/jiot.2016.2640563. URL: http://dx.doi.org/10.1109/jiot.2016.2640563.

[4] Zulfiqar Ahmad Khan et al. "DSPM: Dual sequence prediction model for efficient energy management in micro-grid". In: *Applied Energy* 356 (Feb. 2024), p. 122339. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2023.122339. URL: http://dx.doi.org/10.1016/j.apenergy.2023.122339.

[5] Wendong Zheng and Gang Chen. "An Accurate GRU-Based Power Time-Series Prediction Approach With Selective State Updating and Stochastic Optimization". In: *IEEE Transactions on Cybernetics* 52.12 (Dec. 2022), pp. 13902–13914. ISSN: 2168-2275. DOI: 10.1109/tcyb.2021.3121312. URL: http://dx.doi.org/10.1109/tcyb.2021.3121312.

[6] Zulfiqar Ahmad Khan et al. "Efficient Short-Term Electricity Load Forecasting for Effective Energy Management". In: *Sustainable Energy Technologies and Assessments* 53 (Oct. 2022), p. 102337. ISSN: 2213-1388. DOI: 10.1016/j.seta.2022.102337. URL: http://dx.doi.org/10.1016/j.seta.2022.102337.

[7] Seyed Mohammad Jafar Jalali et al. "A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting". In: *IEEE Transactions on Industrial Informatics* 17.12 (2021), pp. 8243–8253. DOI: 10.1109/TII.2021.3065718.

[8] Lei Tang et al. "Long-term electricity consumption forecasting based on expert prediction and fuzzy Bayesian theory". In: *Energy* 167 (Jan. 2019), pp. 1144–1154. ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.10.073. URL: http://dx.doi.org/10.1016/j.energy.2018.10.073.

[9] S.Sp. Pappas et al. "Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models". In: *Energy* 33.9 (Sept. 2008), pp. 1353–1360. ISSN: 0360-5442. DOI: 10.1016/j.energy.2008.05.008. URL: http://dx.doi.org/10.1016/j.energy.2008.05.008.

[10] Othman Ait Maatallah et al. "Recursive wind speed forecasting based on Hammerstein Auto-Regressive model". In: *Applied Energy* 145 (May 2015), pp. 191–197. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2015.02.032. URL: http://dx.doi.org/10.1016/j.apenergy.2015.02.032.

[11] Zhuang Zheng, Hainan Chen, and Xiaowei Luo. "A Kalman filter-based bottom-up approach for household short-term load forecast". In: *Applied Energy* 250 (Sept. 2019), pp. 882–894. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.05.102. URL: http://dx.doi.org/10.1016/j.apenergy.2019.05.102.

[12] Di Wu et al. "Multiple Kernel Learning-Based Transfer Regression for Electric Load Forecasting". In: *IEEE Transactions on Smart Grid* 11.2 (Mar. 2020), pp. 1183–1192. ISSN: 1949-3061. DOI: 10.1109/tsg.2019.2933413. URL: http://dx.doi.org/10.1109/tsg.2019.2933413.

[13] Zhongfu Tan et al. "Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine". In: *Journal of Cleaner Production* 248 (Mar. 2020), p. 119252. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2019.119252. URL: http://dx.doi.org/10.1016/j.jclepro.2019.119252.

[14] Ping-Huan Kuo and Chiou-Jye Huang. "A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting". In: *Energies* 11.1 (Jan. 2018), p. 213. ISSN: 1996-1073. DOI: 10.3390/en11010213. URL: http://dx.doi.org/10.3390/en11010213.

[15] Mehdi Rafiei et al. "Probabilistic Load Forecasting Using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine". In: *IEEE Transactions on Smart Grid* 9.6 (Nov. 2018), pp. 6961–6971. ISSN: 1949-3061. DOI: 10.1109/tsg.2018.2807845. URL: http://dx.doi.org/10.1109/tsg.2018.2807845.

[16] Jieqiong Zhang. "Optimizing the Analysis of Energy Plants and High-Power Applications Utilizing the Energy Guard Ensemble Selector (EGES)". In: *Informatica* 49.10 (2025), pp. 113–126. DOI: 10.31449/inf.v49i10.7264.

[17] Biaowei Chen et al. "Very-Short-Term Power Prediction for PV Power Plants Using a Simple and Effective RCC-LSTM Model Based on Short Term Multivariate Historical Datasets". In: *Electronics* 9.2 (Feb. 2020), p. 289. ISSN: 2079-9292. DOI: 10.3390/electronics9020289. URL: http://dx.doi.org/10.3390/electronics9020289.

[18] Haixiang Zang et al. "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning". In: *International Journal of Electrical Power; Energy Systems* 118 (June 2020), p. 105790. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2019.105790. URL: http://dx.doi.org/10.1016/j.ijepes.2019.105790.

[19] Yi Zhou et al. "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine". In: *Energy* 204 (Aug. 2020), p. 117894. ISSN: 0360-5442. DOI: 10.1016/j.energy.2020.117894. URL: http://dx.doi.org/10.1016/j.energy.2020.117894.

[20] Lilin Cheng et al. "Multi-Meteorological-Factor-Based Graph Modeling for Photovoltaic Power Forecasting". In: *IEEE Transactions on Sustainable Energy* 12.3 (July 2021), pp. 1593–1603. ISSN: 1949-3037. DOI: 10.1109/tste.2021.3057521. URL: http://dx.doi.org/10.1109/tste.2021.3057521.

[21] Deniz Korkmaz. "SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting". In: *Applied Energy* 300 (Oct. 2021), p. 117410. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.117410. URL: http://dx.doi.org/10.1016/j.apenergy.2021.117410.

[22] Kejun Wang, Xiaoxia Qi, and Hongda Liu. "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network". In: *Applied Energy* 251 (Oct. 2019), p. 113315. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.113315. URL: http://dx.doi.org/10.1016/j.apenergy.2019.113315.

[23] Roozbeh Rajabi and Abouzar Estebsari. "Deep Learning Based Forecasting of Individual Residential Loads Using Recurrence Plots". In: *2019 IEEE Milan PowerTech*. IEEE, June 2019, pp. 1–5. DOI: 10.1109/ptc.2019.8810899. URL: http://dx.doi.org/10.1109/ptc.2019.8810899.

[24] Fath U Min Ullah et al. "Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks". In: *IEEE Access* 8 (2020), pp. 123369–123380. ISSN: 2169-3536. DOI: 10.1109/access.2019.2963045. URL: http://dx.doi.org/10.1109/access.2019.2963045.

[25] Ijaz Ul Haq et al. "Sequential Learning-Based Energy Consumption Prediction Model for Residential and Commercial Sectors". In: *Mathematics* 9.6 (Mar. 2021), p. 605. ISSN: 2227-7390. DOI: 10.3390/math9060605. URL: http://dx.doi.org/10.3390/math9060605.

[26] Tae-Young Kim and Sung-Bae Cho. "Predicting residential energy consumption using CNN-LSTM neural networks". In: *Energy* 182 (Sept. 2019), pp. 72–81. ISSN: 0360-5442. DOI: 10.1016/j.energy.2019.05.230. URL: http://dx.doi.org/10.1016/j.energy.2019.05.230.

[27] Mohamed Abdel-Basset et al. "STLF-Net: Two-stream deep network for short-term load forecasting in residential buildings". In: *Journal of King Saud University - Computer and Information Sciences* 34.7 (July 2022), pp. 4296–4311. ISSN: 1319-1578. DOI: 10.1016/j.jksuci.2022.04.016. URL: http://dx.doi.org/10.1016/j.jksuci.2022.04.016.

[28] Tao Han et al. "An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks". In: *IEEE Internet of Things Journal* 8.5 (Mar. 2021), pp. 3170–3179. ISSN: 2372-2541. DOI: 10.1109/jiot.2020.3013306. URL: http://dx.doi.org/10.1109/jiot.2020.3013306.

[29] Jin-Young Kim and Sung-Bae Cho. "Electric Energy Consumption Prediction by Deep Learning with State Explainable Autoencoder". In: *Energies* 12.4 (Feb. 2019), p. 739. ISSN: 1996-1073. DOI: 10.3390/en12040739. URL: http://dx.doi.org/10.3390/en12040739.

[30] Zulfiqar Khan et al. "Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework". In: *Sensors* 20.5 (Mar. 2020), p. 1399. ISSN: 1424-8220. DOI: 10.3390/s20051399. URL: http://dx.doi.org/10.3390/s20051399.

[31] Gangqiang Li et al. "Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach". In: *Energies* 12.13 (July 2019), p. 2538. ISSN: 1996-1073. DOI: 10.3390/en12132538. URL: http://dx.doi.org/10.3390/en12132538.

[32] Gabriel Trierweiler Ribeiro et al. "Bayesian Optimized Echo State Network Applied to Short-Term Load Forecasting". In: *Energies* 13.9 (May 2020), p. 2390. ISSN: 1996-1073. DOI: 10.3390/en13092390. URL: http://dx.doi.org/10.3390/en13092390.

[33] Shouxiang Wang et al. "Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting". In: *International Journal of Electrical Power; Energy Systems* 109 (July 2019), pp. 470–479. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2019.02.022. URL: http://dx.doi.org/10.1016/j.ijepes.2019.02.022.

[34] Yixing Wang et al. "Short-Term Load Forecasting with Multi-Source Data Using Gated Recurrent Unit Neural Networks". In: *Energies* 11.5 (May 2018), p. 1138. ISSN: 1996-1073. DOI: 10.3390/en11051138. URL: http://dx.doi.org/10.3390/en11051138.

[35] Zulfiqar Ahmad Khan et al. "Efficient Short-Term Electricity Load Forecasting for Effective Energy Management". In: *Sustainable Energy Technologies and Assessments* 53 (Oct. 2022), p. 102337. ISSN: 2213-1388. DOI: 10.1016/j.seta.2022.102337. URL: http://dx.doi.org/10.1016/j.seta.2022.102337.

[36] Meng Zhang. "Load Analysis and Prediction of Electric Vehicle Charging Stations Supported by RF-CNN Algorithm". In: *Informatica* 48.19 (2024), pp. 89–102. DOI: 10.31449/inf.v48i19.6649.

[37] Junhong Kim et al. "Recurrent inception convolution neural network for multi short-term load forecasting". In: *Energy and Buildings* 194 (July 2019), pp. 328–341. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2019.04.034. URL: http://dx.doi.org/10.1016/j.enbuild.2019.04.034.

[38] Noman Khan et al. "AB-Net: A Novel Deep Learning Assisted Framework for Renewable Energy Generation Forecasting". In: *Mathematics* 9.19 (Oct. 2021), p. 2456. ISSN: 2227-7390. DOI: 10.3390/math9192456. URL: http://dx.doi.org/10.3390/math9192456.

[39] Yong Zhang, Yueda Gao, and Zhe Zhao. "Research on Operation and Anomaly Detection of Smart Power Grid Based on Information Technology Using CNN+Bidirectional LSTM". In: *Informatica* 49.7 (2025), pp. 157–164. DOI: 10.31449/inf.v49i7.7037.

[40] Muhammad Sajjad et al. "A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting". In: *IEEE Access* 8 (2020), pp. 143759–143768. ISSN: 2169-3536. DOI: 10.1109/access.2020.3009537. URL: http://dx.doi.org/10.1109/access.2020.3009537.

[41] Fei Wang et al. “Wavelet Decomposition and Convolutional LSTM Networks Based Improved Deep Learning Model for Solar Irradiance Forecasting”. In: *Applied Sciences* 8.8 (Aug. 2018), p. 1286. ISSN: 2076-3417. DOI: 10.3390/app8081286. URL: http://dx.doi.org/10.3390/app8081286.

[42] Jingxuan Liu et al. “A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting”. In: *Applied Energy* 342 (July 2023), p. 121160. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2023.121160. URL: http://dx.doi.org/10.1016/j.apenergy.2023.121160.

[43] Hegang Zong et al. “A Multitask Framework for Optimizing Smart Grid Energy Consumption Using RegClassXNet and Dynamic Cluster Adjustment”. In: *Informatica* 49.18 (2025), pp. 127–142. DOI: 10.31449/inf.v49i18.7865.

[44] Tomás Cabello-López et al. “Forecasting solar energy production in Spain: A comparison of univariate and multivariate models at the national level”. In: *Applied Energy* 350 (Nov. 2023), p. 121645. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2023.121645. URL: http://dx.doi.org/10.1016/j.apenergy.2023.121645.

[45] Chengcheng Jiang and Qunzhi Zhu. “Evaluating the most significant input parameters for forecasting global solar radiation of different sequences based on Informer”. In: *Applied Energy* 348 (Oct. 2023), p. 121544. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2023.121544. URL: http://dx.doi.org/10.1016/j.apenergy.2023.121544.

[46] Ling-Ling Li et al. “Renewable energy prediction: A novel short-term prediction model of photovoltaic output power”. In: *Journal of Cleaner Production* 228 (Aug. 2019), pp. 359–375. ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2019.04.331. URL: http://dx.doi.org/10.1016/j.jclepro.2019.04.331.

[47] Pengtao Li et al. “A hybrid deep learning model for short-term PV power forecasting”. In: *Applied Energy* 259 (Feb. 2020), p. 114216. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.114216. URL: http://dx.doi.org/10.1016/j.apenergy.2019.114216.

[48] Kejun Wang, Xiaoxia Qi, and Hongda Liu. “Photovoltaic power forecasting based LSTM-Convolutional Network”. In: *Energy* 189 (Dec. 2019), p. 116225. ISSN: 0360-5442. DOI: 10.1016/j.energy.2019.116225. URL: http://dx.doi.org/10.1016/j.energy.2019.116225.

[49] Zulfiqar Ahmad Khan et al. “Electrical Energy Prediction in Residential Buildings for Short-Term Horizons Using Hybrid Deep Learning Strategy”. In: *Applied Sciences* 10.23 (Dec. 2020), p. 8634. ISSN: 2076-3417. DOI: 10.3390/app10238634. URL: http://dx.doi.org/10.3390/app10238634.

[50] Zulfiqar Ahmad Khan et al. “A Trapezoid Attention Mechanism for Power Generation and Consumption Forecasting”. In: *IEEE Transactions on Industrial Informatics* 20.4 (Apr. 2024), pp. 5750–5762. ISSN: 1941-0050. DOI: 10.1109/tii.2023.3335453. URL: http://dx.doi.org/10.1109/tii.2023.3335453.