SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I12P116

Volume 11 Issue 12, 170-185, December 2024

Original Article

Enhancing Power Grid Stability through Reactive Power Demand Forecasting Using Deep Learning

Diaa Salman[1*], Abdirahman Farah Ali[1], Suleiman Abdullahi Ali[1], Abdullahi Sheikh Mohamed[1]

[1]*Faculty of Engineering, Jamhuriya University of Science and Technology, Mogadishu, Somalia.*

[1]*Corresponding Author : D.salman@just.edu.so*

Received: 09 October 2024 Revised: 10 November 2024 Accepted: 08 December 2024 Published: 31 December 2024

Abstract *- This paper examines deep learning models for reactive power forecasting in contemporary power systems, emphasizing the Temporal Convolutional Network (TCN), the Transformer, and the models that use both TCN and the Transformer. Reactive power plays a critical role in the stability of power grids, and this paper presents TCN and Transformer models to enhance the forecasting precision of this parameter. Four models, including TCN, Transformer, TCN-Transformer, and Transformer-TCN, were trained and tested based on performance indicators such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Among all the models, the Transformer-TCN hybrid model achieved the best results in all the metrics and successfully learned the short-term and long-term dependencies in the data. The results of the hybrid models were superior to the single models, and the accuracy and stability were improved. The model was validated by checking residual plots and error distributions, and most errors were observed near zero. There were still some prediction challenges in extreme cases, while the Transformer-TCN model was the most accurate for most cases among the architectures analyzed. This work focuses on the prospect of deep learning models for reactive power forecasting and offers a useful tool for managing the power grid. Other research could be directed towards further optimizing hybrid architectures and other methods for enhancing predictive capability in power systems.*

Keywords *- Power systems, Prediction, Deep Learning, Hybrid models, Reactive power.*

1. Introduction

Maintaining the grid's stability is crucial in today's power systems, especially with the increasing complexity and incorporation of renewable energy sources. Another critical element of grid stability is reactive power control, which plays an important role in voltage support across the grid. While active power contributes to doing useful work, reactive power controls the flow of electric currents and the working of electrical appliances. With the rising total amount of energy consumed worldwide and using the percentage of renewable energy sources, it is very important to implement the exact prediction of the reactive power demand to the grid system [1]. The other type of power is reactive power expressed in Volt-Amperes Reactive (VAR), which is crucial in regulating voltage in electrical networks. It helps to maintain the transformers, generators, and other electrical installations within their rated voltage classes. While active power influences the loads, reactive power regulates the voltage, creating the required electromagnetic field in inductive loads like transformers and motors. Low reactive power can cause voltage fluctuations, equipment failure, and even blackouts [2]. In a conventional power system, reactive power control was achieved by using capacitors, inductors, and synchronous condensers. However, with the increasing complexity of modern grids, particularly with the integration of distributed renewable energy resources, the demand for reactive power varies more often. Wind and solar are some renewable energy sources that are unpredictable in terms of their power and, hence, the reactive power required in real-time. Therefore, due to challenges with reactive power demand control, voltage stability has become a significant concern to grid operators, leading to the need for a reliable model for reactive power demand forecasting [3]. The demand for reactive power varies with many factors, such as the load, voltage levels, power output, and even weather conditions, such as temperature and humidity. Unfortunately, these parameters cannot be easily included in a forecasting model because they are highly complex and nonlinear. Conventional system identification techniques, including statistical techniques, simple rules, or predefined tactics, are ineffective when identifying dependent relationships over reactive power. This is made more complex by the fact that forecasting the reactive power is made more difficult by the time-varying relations and nonlinear interaction that the input variables produce [4]. In the past, models such as the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) have been used when forecasting various aspects of power demand. But these models fail when applied to the reactive power demand, especially when it involves AC systems that incorporate the fluctuations of renewable energy sources. Besides, most of these methods are ineffective in maneuvering interaction between two or more variables, which is crucial in modeling [5]. Over the past few years, machine learning and deep learning growth have revolutionized the power system forecasting models. Time series forecasting most usually uses convolutional and recurrent neural networks, including Gated Recurrent Units (GRU) and Long-Short-Term Memory (LSTM). Reactive power forecasting is one of the situations where these models fit since they can handle nonlinear interactions and sequential information [6]. Recent advances in deep learning have presented opportunities for improved reliable and efficient time series forecasting in the power system.

Fully convolutional networks using one-dimensional Temporal Convolutional Networks (TCN) and Transformer models are considered state-of-the-art in forecasting because they capture short- and long-term patterns in the time series data [7]. Temporal Convolutional Networks, or TCNs, are a subcategory of Convolutional Neural Networks (CNNs) designed for sequence analysis. Unlike the conventional RNNs, TCNs apply causal convolutions to ensure that the output of a specific time step depends on the inputs that have occurred before that time step, which is very important for time series forecasting. Another major benefit of TCNs over other RNN-based models like LSTM and GRU is that TCNs do not face the problem of vanishing gradients in long sequences.

The TCNs can effectively model both local and long-range dependencies by utilizing dilated convolutions. The network can look back over long periods without incurring a high computational cost [8]. Besides TCNs, Transformer models have also outperformed other models in different time-series forecasting problems. Transformers were first used in natural language processing applications but are now used in power system forecasting because of their capability to model the dependencies between time steps. Compared with the traditional RNNs, the Transformers can self-learn the attention of all the time steps simultaneously, making it easier for the model to learn the correlation between the distant time steps. This makes Transformers particularly suitable for applications like reactive power forecasting, where the long-term dependencies and interdependencies between the variables must be considered [9].

The self-attention mechanism in Transformers helps the model assign different importance levels to different time steps, which is not the case with the standard sequential models. This is particularly significant in reactive power demand forecasting since some of these factors (e.g., voltage fluctuations) may take a long to present their effects. While TCN and Transformer models have been previously implemented to achieve favorable results in time-series forecasting, further integrating the two into a single model could provide even better results. Combined models utilize each model's advantages and can more accurately represent local and global dependencies in data. In this paper, we try to apply TCN-Transformer models for forecasting the reactive power demand, which can learn local patterns of TCNs and long-range sequence modeling of transformers [10, 11]. First, by using the TCN layers, the model can learn short-term dependencies and local connections between features such as voltage levels and system load. The Transformer layers can then operate on the output of the TCN to capture more complex, long-term dependencies across the whole time series. This mixed model method enables the model to predict short-term and long-term oscillation trends for the reactive power demand forecasting of highly volatile grid systems.

Even though TCN and Transformer models have been used in many fields, including speech recognition, language translation, and financial time-series forecasting, their application in power system forecasting, especially in the forecasting of reactive power demand, is still rare. Some of the previous works that have been done are on active power demand forecasting using TCNs and Transformers. For example, Almqvist et al. proved that using a TCN-based model improves the active power demand forecasting in renewable energy systems compared to traditional ARIMA models [12]. Likewise, Sun et al. used a Transformer-based model to predict wind power generation. They proved the model could identify short-term and long-term trends in wind power generation capability [13]. However, although there have been studies on the use of TCN and Transformer models for active power forecasting, there have been few studies on using these models for reactive power demand forecasting. Most of the previous works have considered active power forecasting. At the same time, few of them have addressed using these sophisticated models to predict the reactive power demand, which is more challenging and significant than active power and depends much on the voltage stability.

To this end, this paper seeks to apply and assess TCN-Transformer hybrid models for reactive power demand forecasting. Based on these considerations, we conjecture that TCN's ability to learn local patterns and Transformer's capacity to model long-term dependencies will improve forecasting accuracy and resilience in highly volatile grid contexts where the demand for reactive power is known to change frequently. The main aim of this paper is to analyze the effectiveness of TCN, Transformer, and TCN-Transformer models in predicting reactive power demand. In this paper, we compare these models' results to determine if the hybrid models are superior to the standalone deep learning models in capturing the temporal dependencies that affect reactive power demand. Specifically, we emphasize using these models to address the nonlinear and time-varying characteristics of the reactive power variation in the modern power system.

2. Literature Review

In the last few years, the increasing demand for more efficient grid operations and the rising complexity of systems have led to substantial attention on integrating machine learning techniques. Reactive power forecasting is especially important among numerous forecasting tasks since it helps maintain voltage stability and improve the efficiency of power system operations. For this reason, various deep learning models have been investigated in the literature, with significant progress in adopting TCNs, Transformers, and hybrid designs that fuse different models.

2.1. Temporal Convolutional Networks (TCN) in Power Forecasting

Especially in the framework of power systems, TCNs have emerged as a strong substitute for conventional Recurrent Neural Networks (RNNs) for problems involving time series forecasting. Unlike RNNs that process data sequentially, TCNs efficiently model both short-term and long-term dependencies inside time-series data using causal convolutions combined with dilated filters. Results show that TCNs are useful for several forecasting tasks, including load forecasting in line with wind power projection. Bai et al. (2018) showed that TCNs are superior to LSTMs in a variety of sequence modeling tasks, pointing out their higher efficiency in capturing temporal dependencies without the gradient vanishing issues typical of RNNs and LSTMs [14], while Yu et al. (2019) investigated TCNs for short-term load forecasting, reporting that they outperformed other deep learning models in both accuracies and compute ability of the TCN to handle the temporal characteristics of power system variables suggests that it is a feasible choice for reactive power forecasting. While TCNs perform admirably in short-term dependency capture, they may have difficulty handling complex long-range dependencies, which is vital for reactive power forecasting because of historical patterns.

2.2. Transformer Models in Power Systems

The Transformer model, first presented for Natural Language Processing (NLP) by Vaswani et al. (2017), has become popular in time-series forecasting owing to its talent for modeling long-term relationships using self-attention mechanisms [15]. Transformers, unlike RNNs, handle data in parallel thanks to their self-attention mechanisms, which improve scalability. This has rendered them very effective in depicting long-range dependencies, an important need in tasks such as reactive power forecasting, which involves considering interactions between various system variables across time. Lai et al. (2019) recently showed the utility of Transformers for load and demand forecasting, pointing out that the attention mechanism helps the model to hone in on significant time steps, which improves prediction accuracy [16]. Similarly, Zhou et al. applied Transformers to multivariate time series forecasting in electricity consumption. They observed that the model's skill in capturing both global and local dependencies improved the forecasting performance. These studies underline the ability of Trancaps to deal with the complexities of energy systems, where forecasting reactive power involves critical long-term dependencies and interactions among voltage, current, and active power. A shortcoming of standalone Transformer models is their propensity to need large volumes of data for training and possible issues in encapsulating local temporal patterns. This limitation has resulted in the investigation of hybrid models exploiting several architectures' strengths.

2.3. Existing Gaps in Reactive Power Forecasting

Load and renewable energy forecasting have received much attention in research, but reactive power forecasting has not yet been much-concentrated. Much of the literature currently highlights active power, demand forecasting, or renewable energy generation, often giving reactive power a minor role. Given the important role reactive power plays in supporting voltage stability and optimizing power system performance, there is a major gap in the existing literature. Tolun and Zor (2024) studied the importance of improving reactive power forecasting accuracy with machine learning algorithms, including LSTM, GRU, and XGBoost [17].

This study pointed out the crucial role of reactive power in the stability of modern energy systems, particularly in locations such as hospitals where uninterrupted energy supply and voltage stability are critical. The investigation pointed out the difficulties of predicting reactive power because of its nonlinear character and dependence on different electrical and meteorological factors. Focusing on very short-term reactive power forecasting within a large hospital complex, the study delivered essential insights into the performance of machine learning models in this field. A limited body of research into reactive power forecasting, especially with advanced models like Transformers and hybrid architectures, points to the requirement for more detailed studies. This work aims to bridge this gap by applying both hybrid TCN-Transformer and Transformer-TCN models to the problem of reactive power forecasting while assessing how well they capture both short-term and long-term dependencies. This research can improve both voltage stability and the reliability of the entire grid.

3. Methodology

3.1. Temporal Convolutional Network

The Temporal Convolutional Network (TCN) is a deep learning architecture for sequence modeling and time-series forecasting. Unlike other recurrent neural networks commonly used for sequential data, TCNs employ convolutional layers while preserving the sequential order of inputs. This makes TCNs highly efficient, allowing them to learn both local and long-range dependencies in the data, and they have none of the vanishing gradient problems associated with other recurrent models like LSTM and GRU. The major concept of TCN is to use convolutional filters on temporal sequences. This makes it possible for the model to capture intricate relationships after several time steps. Unlike the sequential operations used in

RNNs, TCNs use one-dimensional convolutions across the inputs to simultaneously process all the time steps. This leads to shorter training times and the ability to capture temporal dependencies. Another advantage of TCN is that it employs causal convolutions, which means that the predictions at the time step depend only on the preceding time steps (i.e., past information). This makes TCNs a powerful tool for such tasks as a forecast, in which the model has to predict future values relying only on historical data. Another interesting component is dilated convolution, which allows the model to obtain more extended context information without adding extra computations. Dilations make it possible to look back in time over larger periods by jumping some time steps in the sequence, thus allowing the model to consider information from both short and long-time steps. For reactive power forecasting, TCNs can be applied to represent the temporal dependencies between the active power, voltage levels, and the reactive power demand, as illustrated in Figure 1. This section describes the mathematical conceptualization of the TCN by using symbols pertinent to the issue of reactive power forecasting [18].

$$\mathrm{X}(t) = [P(t-n), V(t-n), I(t-n)], \dots, [P(t), V(t), I(t)] \tag{1}$$

Where X(t) represents the input sequence at time t, P(t), V(t), and I(t) represent the active power and voltage current intensity at time t, respectively. In TCNs, the causal convolution ensures that the output at time step t, denoted as Q ˆ(t), is only influenced by past values X(t), X(t-1),…, X(t-n). Mathematically, the causal convolution is represented as:

$$\hat{\mathrm{Q}}(\mathrm{t}) = \sum_{\mathrm{k}=0}^{\mathrm{K}-1} \mathrm{W}_{\mathrm{k}} \cdot \mathrm{X}(\mathrm{t}-\mathrm{k}) \tag{2}$$

Where K is the size of the convolutional filter, W_{k} is the filter weight at the k-th position, and $\mathrm{X}(\mathrm{t}-\mathrm{k})$ is the input vector at time $\mathrm{t}-\mathrm{k}$ that includes $\mathrm{P}(\mathrm{t}-\mathrm{n}), \mathrm{V}(\mathrm{t}-\mathrm{n}), \mathrm{I}(\mathrm{t}-\mathrm{n})$.

This keeps the model blind without using values of getting points at time steps more than t, a property referred to as causal convolution. Since temporal dependencies extend over longer periods, the size of the convolutional filters cannot be enlarged, and thus, dilated convolutions are used. A dilated convolution is a standard convolution with spaces or what is referred to as the dilation factor between the filter elements. This allows the TCN to consider inputs spaced out in time, yet the complexity of the model does not increase as follows.

$$\hat{Q}(t) = \sum_{k=0}^{K-1} \mathrm{W}_k \cdot \mathrm{X}(t-d \cdot k) \tag{3}$$

D is the dilation factor, which controls the spacing between the filter elements, and $\mathrm{X}(\mathrm{t}-\mathrm{d} \cdot \mathrm{k})$ is the input vector at a delayed time step. An important component of the TCN's architecture is the application of residual connections, which allows deep network training and solves the problem of vanishing gradients. In the residual block, the input is added directly to the result of the convolutions. Hence, the network can learn the residuals and the difference between the input, output, and actual outputs. The residual connection is mathematically represented as:

$$\hat{\mathrm{Q}}(\mathrm{t}) = \mathrm{X}(\mathrm{t}) + \mathrm{f}(\mathrm{X}(\mathrm{t}), \mathrm{W}) \tag{4}$$

Where $\mathrm{f}(\mathrm{X}(\mathrm{t}), \mathrm{W})$ represents the series of a convolution applied to the input sequence $\mathrm{X}(\mathrm{t})$.

$$\hat{\mathrm{Q}}(\mathrm{t}) = \mathrm{W}_{\mathrm{out}}.\mathrm{h}(\mathrm{t}) + \mathrm{b} \tag{5}$$

Where, $\mathrm{h}(\mathrm{t})$ is the output of the final convolutional layer, $\mathrm{W}_{\mathrm{out}}$ are the output layer weights, and b is the bias term.

3.2. Transformer Model

The Transformer model is one of the deep learning architectures Google proposed in a paper by Vaswani et al. (2017). Initially, the Transformer was developed for natural language processing tasks but has been applied to other problems, such as time series forecasting, because of the model's ability to capture long-range dependencies in the data. In contrast to the recurrent neural networks that work sequentially, Transformers use self-attention, allowing the model to learn dependencies and process them simultaneously.

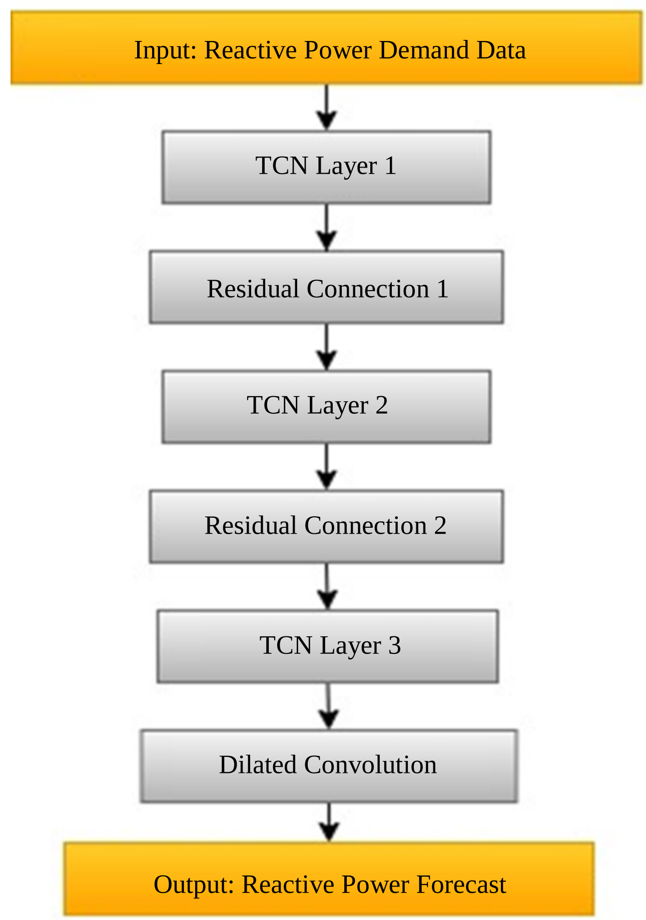

Fig. 1 TCN model structure

The self-attention mechanism can attend to different parts of the input sequence when predicting, as illustrated in Figure 2. This is very relevant in the case of reactive power forecasting, where the future values of Q(t) are conditioned not only by past values but also by other factors like P(t), V(t), and I(t) [15].

The Transformer model's mathematical representation can be simplified by two equations while focusing on its key components for reactive power forecasting as follows:

$$\mathrm{A}(t) = \sum_{t'=1}^{n} \operatorname{softmax}\left(\frac{\mathrm{Q}(t) \cdot \mathrm{K}(t')}{\sqrt{d_k}}\right) \cdot \mathrm{V}(t') \tag{6}$$

Where, $\mathrm{Q(t)}, \mathrm{K(t)},$ and $\mathrm{V(t)}$ are the query, key, and value vectors at time t, calculated as a linear transformation of the input sequence, $\mathrm{d_k}$ is the dimensionality of the key vectors (used for scaling), and the attention score determines how much importance is given to other time steps $\mathrm{t'}$ in the input sequence when predicting time t.

The final prediction (Using the Output of the Attention Mechanism) can be represented as follows:

$$\hat{Q}(t) = \mathrm{W}_o \cdot \mathrm{O}(t) + b_o \tag{7}$$

Where $\mathrm{O(t)}$ is the output of the Transformer encoder after passing through the attention and feed-forward layers, $\mathrm{W_o}$ and $\mathrm{b_o}$ are the weights and bias of the final output layer and $\mathrm{\hat{Q}(t)}$ is the predicted reactive power at time t.

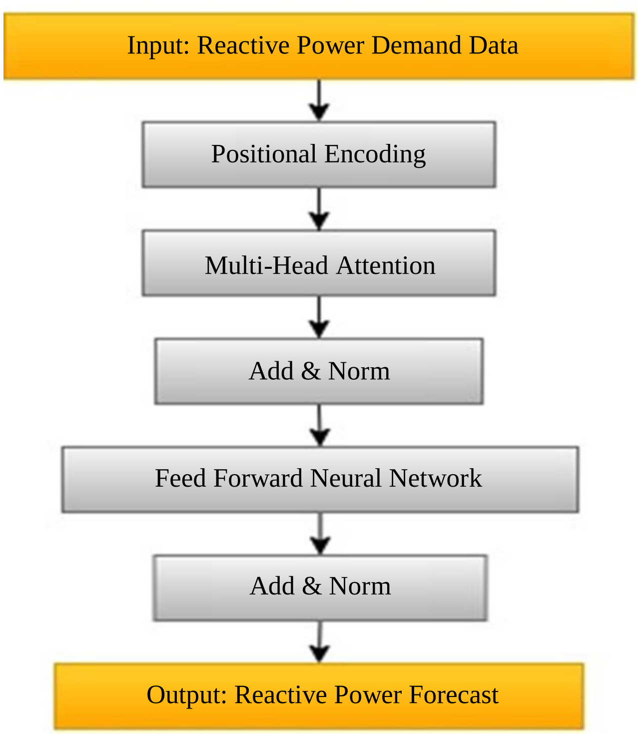

Fig. 2 Transformer model structure

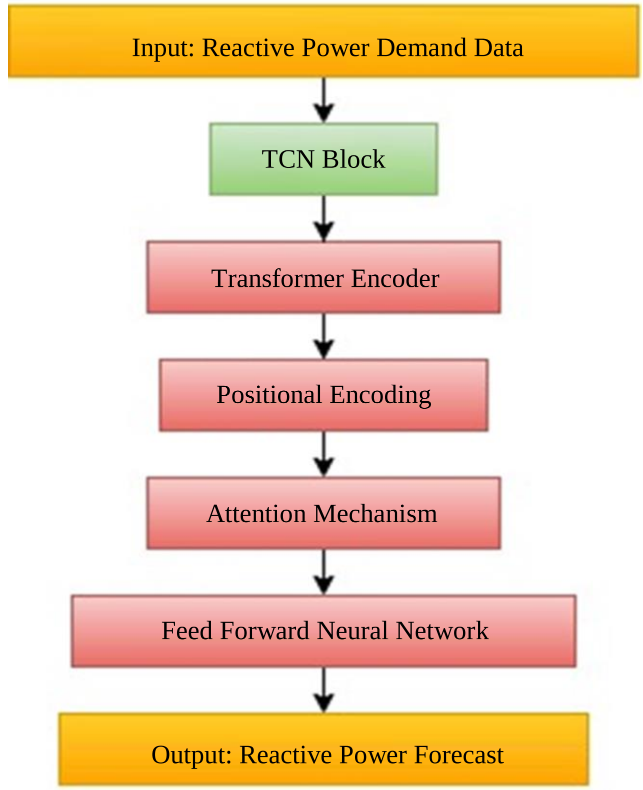

Fig. 3 TCN-transformer model structure

3.3. Hybrid Models

Recently, the use of hybrid models has been considered in time-series forecasting because of their potential to integrate the characteristics of various deep-learning architectures. In the case of hybrid TCN-transformer and transformer-TCN models, the concept combines the benefits of both TCN and transformer models. These models are intended to capture different aspects of the temporal dynamics, making it possible to make a more accurate prediction of reactive power Q(t).

3.3.1. Hybrid TCN-Transformer

The hybrid TCN-Transformer model combines the TCN and the Transformer model, where the input sequence is first fed to the TCN, which works on the temporal patterns with the help of causal and dilated convolution. Then, the output of TCN is passed to the Transformer, which learns the long-range dependencies by its self-attention mechanism [19].

The concept is that the TCN effectively captures microscale features (voltage oscillations or short-term fluctuations in active power). At the same time, the Transformer improves the model's capacity to model intertemporal dependencies across time steps for long-term reactive power forecasting, as illustrated in Figure 3. The mathematical representation of the hybrid model can be summarized as follows:

Let $\mathrm{H}_{\mathrm{TCN}}(\mathrm{t})$ represent the output of the TCN block at time t :

$$\mathrm{H}_{TCN}(t) = f_{TCN}(\mathrm{X}(t-n), \ldots, \mathrm{X}(t)) \quad (8)$$

Where $\mathrm{f}_{\mathrm{TCN}}$ represents the series of causal and dilated convolutions applied by the TCN.

The output from the TCN block is then passed into the Transformer encoder. The self-attention mechanism in the Transformer operates on $\mathrm{H}_{\mathrm{TCN}}(\mathrm{t})$ as follows:

$$\mathrm{A}(t) = \sum_{t'=1}^{n} \operatorname{softmax}\left(\frac{\mathrm{Q}(t) \cdot \mathrm{K}(t')}{\sqrt{d_k}}\right) \cdot \mathrm{V}(t') \quad (9)$$

Where $\mathrm{Q}(\mathrm{t}), \mathrm{K}(\mathrm{t}), \mathrm{V}(\mathrm{t})$ are the query, key, and value vectors derived from $\mathrm{H}_{\mathrm{TCN}}(\mathrm{t})$.

Finally, the output of the attention mechanism $\mathrm{A}(\mathrm{t})$ is used to predict the reactive power $\hat{\mathrm{Q}}(\mathrm{t})$:

$$\hat{Q}(t) = \mathrm{W}_o \cdot \mathrm{A}(t) + b_o \quad (10)$$

3.3.2. Hybrid Transformer-TCN

The Hybrid Transformer-TCN is the model whose architecture is opposite to the previous one, and after applying the Transformer layer, the output is passed through the TCN block. The Transformer first encodes the dependencies between the distant positions and determines which time steps are relevant to predicting reactive power. Subsequently, the TCN is employed to polish the sequence based on its causal and dilated convection to identify local temporal features that the Transformer may overlook.

The input sequence is first fed into the Transformer encoder block. The self-attention mechanism helps to find relevant time steps across the sequence and produces a better representation of the sequence, considering short-term and long-term dependencies. After that, the output of the Transformer encoder is fed to the TCN, which involves causal and dilated convolutions to process the sequence to maintain temporal causality in the model, meaning that no following time steps can influence previous predictions while capturing local patterns as illustrated in Figure 4. The mathematical representation of the hybrid model can be summarized as follows:

Let $\mathrm{H}_{\text{Transformer}}(\mathrm{t})$ represent the output of the Transformer encoder at time t :

$$\mathrm{H}_{\text{Transformer}}(t) = f_{\text{Transformer}}(\mathrm{X}(t-n), \ldots, \mathrm{X}(t)) \quad (11)$$

Where $\mathrm{f}_{\text{Transformer}}$ represents the self-attention mechanism applied by the Transformer. The output from the Transformer is then passed into the TCN block. The TCN processes the sequence using dilated convolutions:

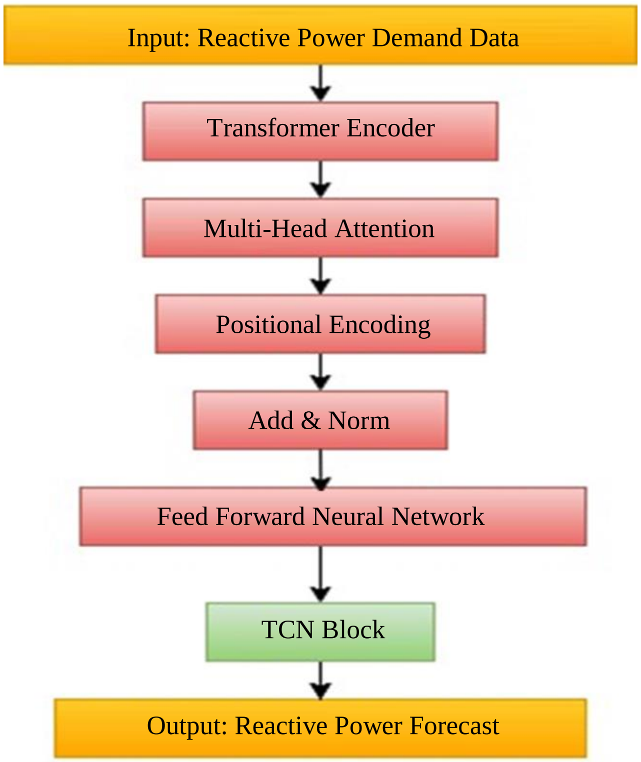

Fig. 4 Transformer-TCN model structure

$$\mathrm{H}_{TCN}(t) = \sum_{k=0}^{K-1} \mathrm{W}_k \cdot \mathrm{H}_{\text{Transformer}}(t - d \cdot k) \quad (12)$$

Where K is the kernel size, W_{k} are the filter weights, and d is the dilation rate.

Finally, the output of the TCN $\mathrm{H}_{\mathrm{TCN}}(\mathrm{t})$ is used to predict the reactive power $\hat{\mathrm{Q}}(\mathrm{t})$:

$$\hat{Q}(t) = \mathrm{W}_o \cdot \mathrm{H}_{TCN}(t) + b_o \quad (13)$$

The Hybrid TCN-Transformer and Hybrid Transformer-TCN models seek to integrate the TCN and Transformer structures to obtain enhanced predictive accuracy in reactive power forecasting. While the TCN-Transformer learns localized patterns first, the Transformer-TCN learns global dependencies first and then processes the sequence with causal convolutions. Each hybrid model has its benefits, and the performance of the two models in terms of capturing short and long-term dependencies in the power system data can be compared.

4. Data Description

The data used in this study are collected from the public domain energy consumption data, which includes household and system electrical parameters. It also involves time series data that describes the dynamic of active and reactive power consumption over time, making it more appropriate for the relevant forecasting activities in this study.

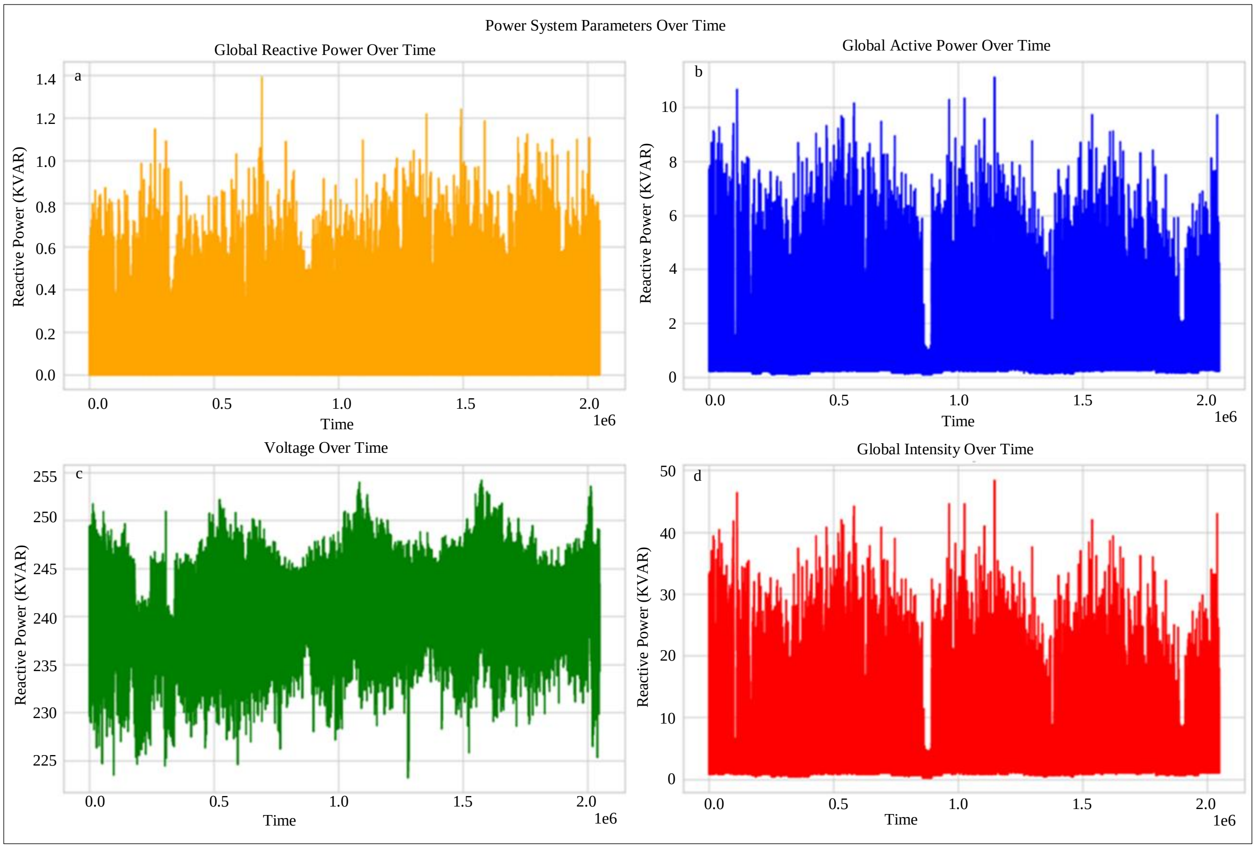

Fig. 5 Time series analysis of key power system parameters: (a) Global reactive power (b) Global active power, (c) Voltage (d) Global intensity

The data was collected at regular one-minute intervals, giving a good picture of power demand, voltage levels, and current intensity in a power grid [20]. Figure 5 illustrates the feature's performance over time. The variables of interest in the dataset are Global_active_power and Global_reactive_power, expressed in kilowatts (kW) and kilovolt amperes reactive (kVAR), respectively. Global_active_power is the real power consumed by the system and can-do work such as running appliances or turning on the light. Global_reactive_power, in contrast, is the reactive power, which is the power that fluctuates between the source and the load and is important in regulating voltage in the power network.

Therefore, reactive power demand forecasting is a key focus of this study since it is a key factor in maintaining grid stability. Besides, the dataset contains a Voltage variable in volts that reflects the electrical potential difference at the point of measurement and Global_intensity in amperes as the total current consumed by the system. These variables provide important background information on power consumption behaviour and are useful for modeling reactive power demand. Additionally, three sub-metering measurements, Sub_metering_1, Sub_metering_2, and Sub_metering_3, represent active energy consumed by circuits or sections of the household system in watt-hours (Wh). Although these sub-metering values mainly reflect certain energy consumption trends in the system, they are very informative about power consumption distribution. Generally, the dataset offers a comprehensive picture of active and reactive power consumption profiles over time. It is well suited to training and testing deep learning models for predicting reactive power demand. Such a detailed understanding of the power consumption patterns is crucial for controlling the power grid stability and effectively controlling the reactive power flows.

From the heatmap of Figure 6, it is easy to infer the correlation between the various electrical parameters in the dataset. It shows how some variables are positively related while others are weak or negatively related. For instance, the Global_active_power variable has a very high positive correlation with the Global_intensity variable; this means that as the active power consumed by the system increases, so does the current intensity. This is expected because, in electrical systems, power is directly proportional to current. Further, it is observed that the Global_active_power is highly correlated with one of the sub-metering units, Sub_metering_3, which suggests that the devices or circuits that this sub-meter measures are responsible for a large part of the total power consumption.

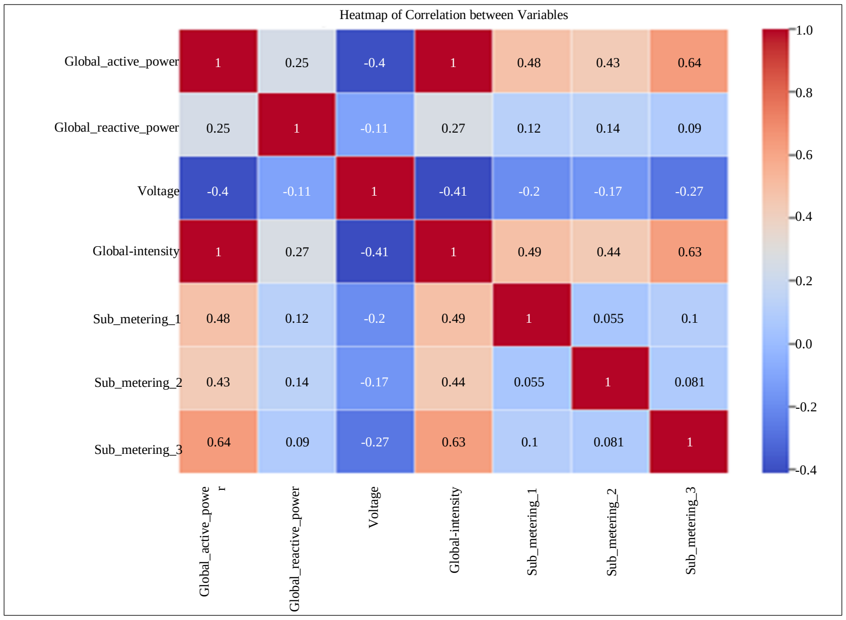

Fig. 6 Correlation heatmap of power system parameters

In turn, Global_reactive_power has relatively low coefficients of determination with most of the other variables, including active power and intensity, which indicates that the demand for reactive power is somewhat independent of the other parameters, at least in the framework of this dataset. The voltage in the system has an inverse relationship with both the active power and the intensity; in other words, the lower the voltage, the higher the power consumption and current density. This aligns with normal power system characteristics, whereby lower voltage levels result in higher currents for the same power consumption. Figure 7 is a pair plot visually representing the correlation between the two power system variables: Global_reactive_power, Global_active_power, Voltage, and Global_intensity. Each plot represents the relationship between two variables, and the diagonal represents the kernel density estimate for each variable. Several trends can be deduced from the pair plot. First, there is a direct correlation between Global_active_power and Global_intensity, as the scatter plot of the two variables shows a diagonal line. This agrees with the relationship between active power and current normally observed in electrical systems. On the other hand, Global_reactive_power does not have a very high correlation with the other variables in the form of a straight line. This is evidenced by the scatter plots between Global_reactive_power and the other variables, such as Global_active_power and Voltage, which are more spread out, implying a weaker correlation. This implies that other factors not captured in this dataset might affect reactive power. Voltage distribution seems fairly normal, as indicated by the bell-shaped density plot along the diagonal. Global_reactive_power and Global_intensity, on the other hand, are more skewed, with a higher number of data points located at the lower end of the scale and fewer at the higher end.

5. Research Workflow

To establish a well-structured method of forecasting reactive power using different deep learning models such as TCN, Transformer, and Hybrid architectures, the following workflow is adopted and illustrated in a flowchart as in Figure 8. Data cleaning and preprocessing form the first step of the process, wherein the data collected is prepared for model training. This entails dealing with missing data, excluding outliers, and making structured sequences from available variables, including active power, voltage, and current intensity. This is where the scaling of the data is also done to normalize the features so that it can be easy for the models to find patterns in the data. Further, sequences are generated to transform the dataset into the time steps needed for the temporal models.

After data cleaning and preparation, the next step is model selection. This involves the selection of architectures, which include TCN and Transformer, and the combined models, which include TCN-Transformer and Transformer-TCN. The choice of the model may include the nature of the data, the understanding of the system's behavior, or a trial-and-error approach as to which architecture works best. Each model has its strength: TCN is intended to model short-term temporal dependencies, while the Transformer can model long-term dependencies through an attention mechanism. After model selection, the data is divided into the training and test data sets. The training set is used to learn the patterns in the data, and the testing set is used to test the model's performance on unseen data. The split helps to check whether the model is capable of generalizing to other data or not. Special attention is paid to the fact that the training and testing sets are chosen from the entire data set, and the data's temporal structure is considered.

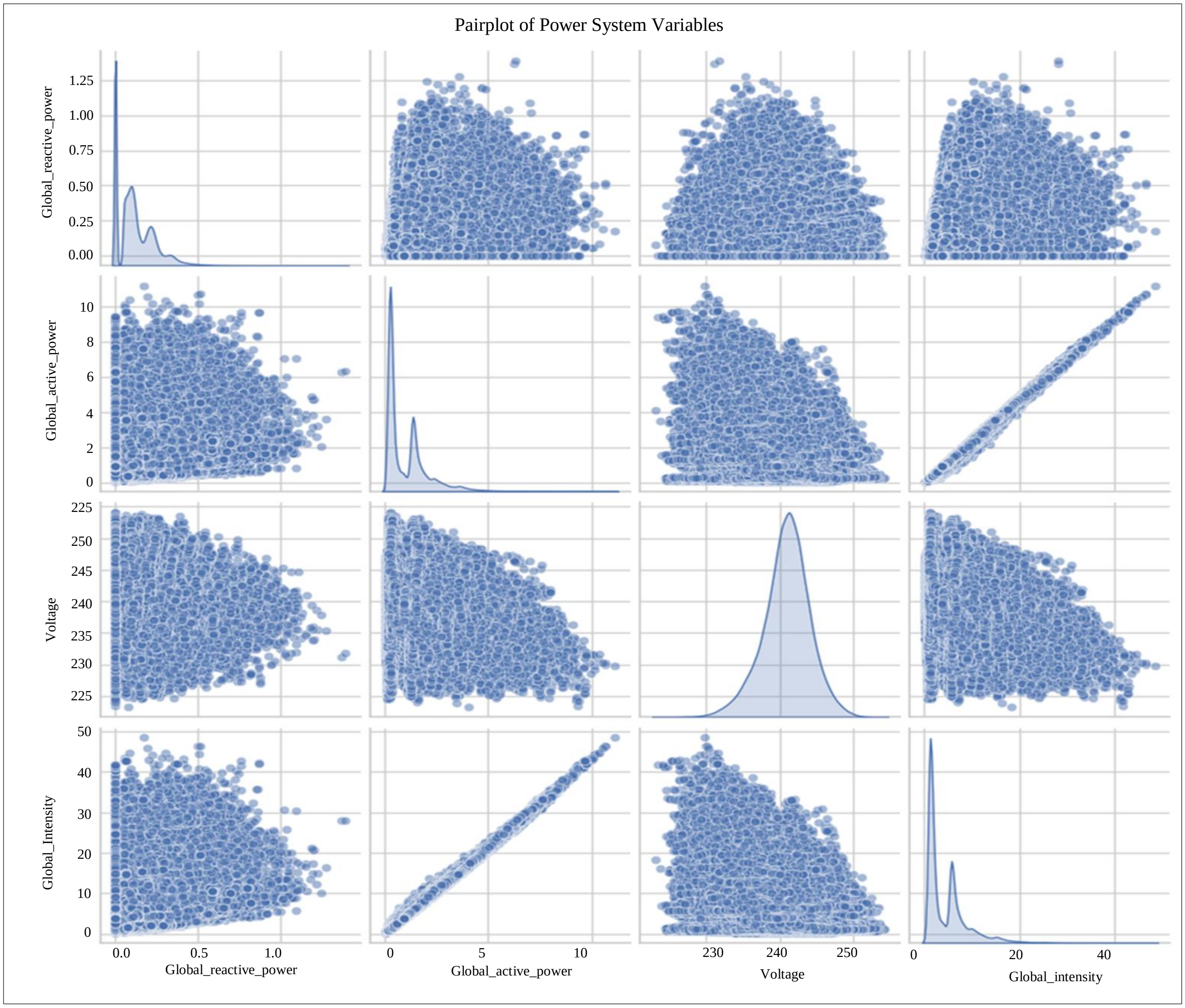

Fig. 7 Pairwise relationships of power system variables

Then, the selected model is trained and tested using the split data. In the training phase, the model learns the mapping between the input features, for instance, active power and voltage, and the target variable, reactive power. The optimization algorithm then adjusts the model to minimize a loss function, for example, Mean Squared Error (MSE). After model training, the model's prediction is tested on the testing set. This phase involves the application of the model to forecast reactive power and then compare it with actual values. The next step is performance evaluation, where it is possible to calculate different performance measures, including Mean Absolute Error (MAE), MSE, and Root Mean Squared Error (RMSE). These metrics give a quantitative evaluation of the model's ability to predict the reactive power. This way, the metrics of the different models are compared, and it is possible to see which architecture best suits the task. This step is important in establishing the ability of the model to capture the right patterns in the data and its ability to generalize on unseen data.

After assessing the model's performance, the best model is chosen depending on the calculated values. If the model has achieved the required operational characteristics, it is considered complete and proceeds to the next step. However, suppose the performance of this model is not satisfactory. In that case, the workflow may be backtracked to the model selection step to try out other models or adjust the current model's parameters. Last of all, the performance visualization is done to show how well the model predicts the reactive power. Different plots can be employed: the line plot of the predicted and actual values, the scatter plot of the predicted and actual reactive power, and the residual plot for the errors. These are very useful, especially in interpreting the model and identifying areas that may require fine-tuning.

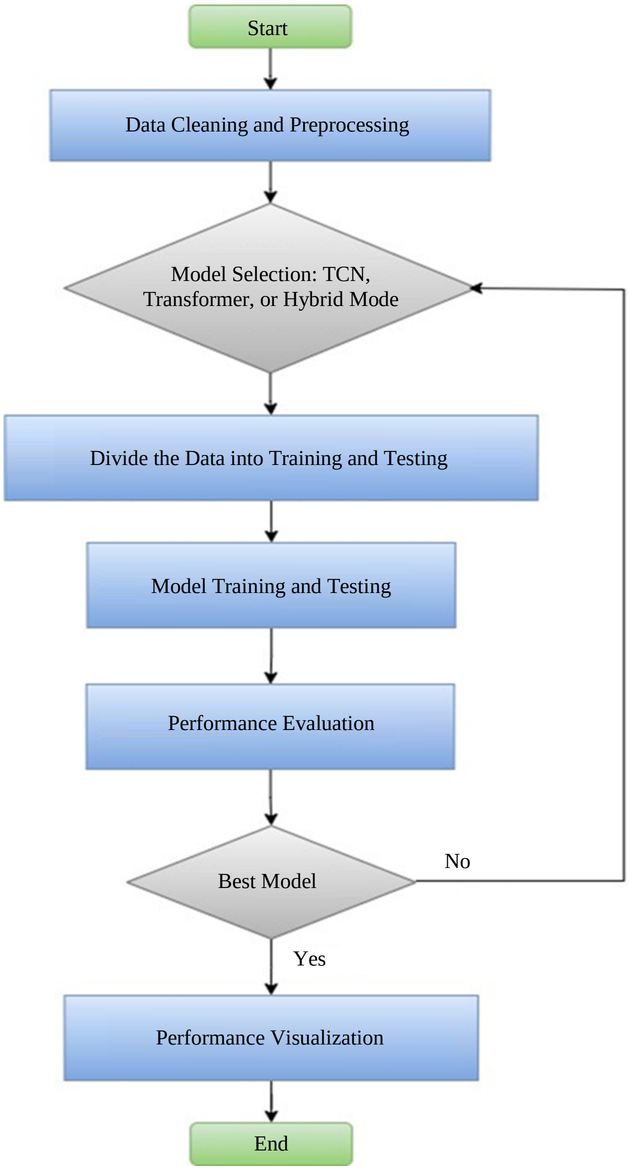

Fig. 8 Reactive power demand forecasting workflow

Table 1. A comparison between the model's performance

Model	MAE	MSE	RMSE
TCN	0.064125	0.008214	0.090632
Transformer	0.064401	0.007636	0.087385
TCN-Transformer	0.067017	0.007301	0.085443
Transformer-TCN	0.059869	0.006783	0.082364

6. Results and Discussions

Table 1 provides a comparative analysis of four different deep learning models—TCN, Transformer, TCN-Transformer, and Transformer-TCN evaluated based on three performance metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These metrics are crucial for evaluating the accuracy and variability of the predictions in reactive power forecasting. The MAE stands for Mean Absolute Error; it reveals the mean of the absolute differences between the predicted and actual values without the direction for the errors. The proposed Transformer-TCN model has the lowest MAE, which confirms that it provides the best approximation of the real reactive power values. The TCN and Transformer models give a slightly higher but comparable MAE, which makes them rank right behind it. On the other hand, the TCN-Transformer model yields the highest MAE, meaning that the model's predictive capability is slightly poorer than the other models. When it comes to MSE, which gives larger errors more weight by squaring the prediction errors, the Transformer-TCN model has the lowest score, further proving its efficiency in reducing significant prediction errors. This is important, especially because it indicates the model's resistance to noise, as it is in statistics, outliers, or anomalous data. The TCN-Transformer model has the highest MAE but has a lower MSE than both the individual Transformer and TCN models, which suggests that it may be less volatile in larger errors than when used independently.

The RMSE, which is the square root of MSE and offers the error metrics in the same units as the reactive power, also corroborates the previous observations. The Transformer-TCN model presents the lowest RMSE, which indicates the model's highest and most stable performance for all the predictions. The TCN model, while performing relatively well in other aspects, has the highest RMSE, suggesting its relatively higher variation and variability in error size. The Transformer-TCN model generally has the lowest performance in all three metrics, while it is superior to the other models in each case. This hybrid model allows the Transformer to model global dependency while benefiting from TCN's local pattern recognition ability to forecast reactive power. The proposed Transformer-TCN model records a better performance and substantiates the effectiveness of the hybrid models for intricate forecast problems by integrating the features of various architectures to improve the prediction efficacy and robustness. On the other hand, the TCN and Transformer models, while good on their own, are slightly below the performance of the hybrid models, which suggests that a single-model approach may not be quite adequate for the complexities of reactive power forecasting.

The bar chart of Figure 9 provides a visual comparison of the four models—TCN, Transformer, TCN-Transformer, and Transformer-TCN based on their performance across three key metrics: MAE, MSE, and RMSE. The blue bars refer to the MAE, the green bars refer to the MSE, and the orange bars refer to the RMSE. The proposed Transformer-TCN model outperforms all the other models in all the metrics, as shown by the lowest scores. On the other hand, TCN and TCN-Transformer models present higher error rates, especially RMSE, as the difference between the prediction and the real values.

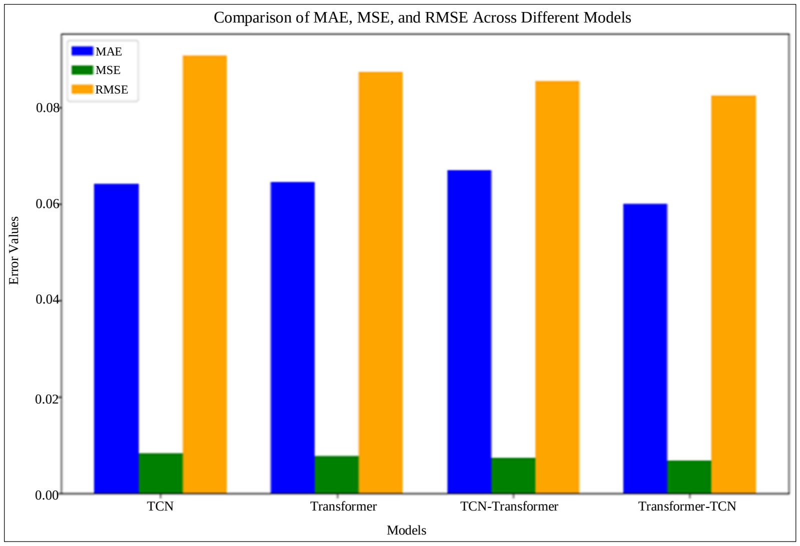

Fig. 9 A comparison between the model's performance

The chart corroborates the hypothesis supporting hybrid models, as the Transformer-TCN structure is the most accurate and consistent in estimating reactive power. In the case of the Transformer-TCN model, Figure 10 shows the training and validation loss in MSE for 200 epochs. The blue line refers to the training loss, which consistently decreases as the model trains, suggesting that the model effectively minimizes the training error. The orange line, the validation loss, decreases for the first five epochs and then has a more erratic trend than the training loss. As the validation loss starts to level off at some fixed value, it is always higher than the training loss in the process. This could be a sign of slight overfitting, where the model is great at fitting the training data but not as good at applying those same methods to the unseen validation data. However, the overall loss trends downward, indicating that the model is learning. However, there may be better ways to generalize, potentially through future tweaks, training, and regularization or more accurate early stopping. Using the Transformer-TCN hybrid model, Figure 11 presents the True and Predicted Global Reactive Power during the first 100 samples. The blue line shows reactive power values, while the orange line shows the reactive power values estimated by the model. It can be seen that the two lines are maintained almost parallel to each other for most of the samples, proving the efficacy of the model in mapping the basic characteristics of the reactive power demand reaction.

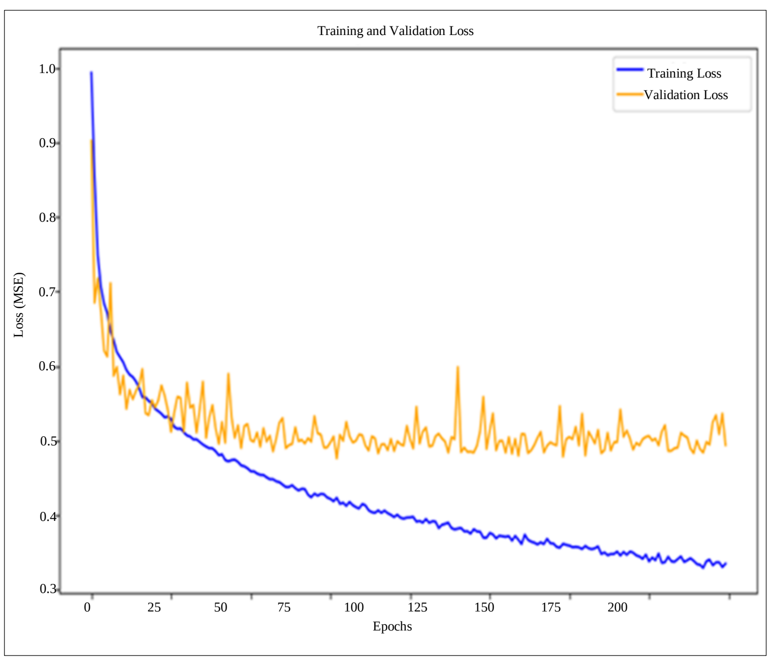

Fig. 10 The training and validation loss of the transformer-TCN model

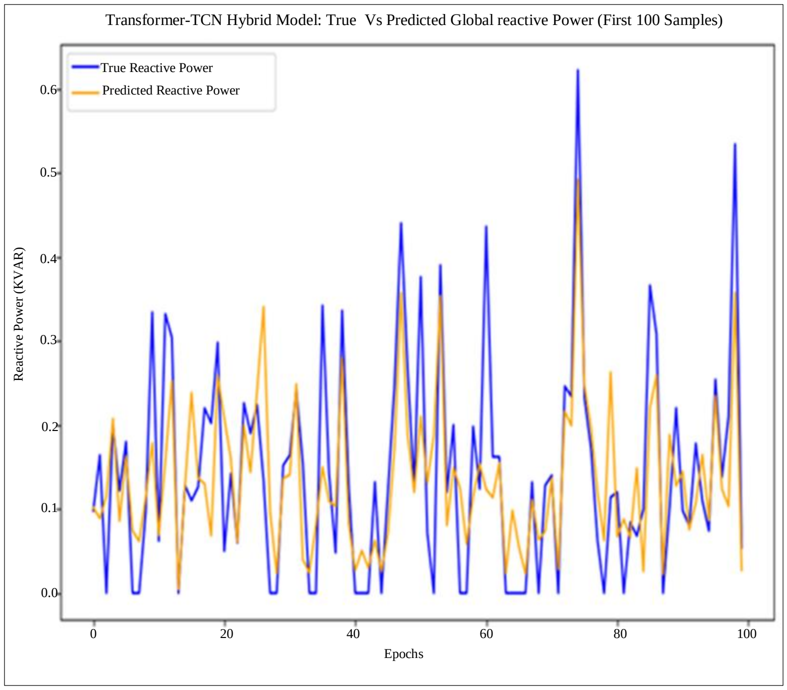

Fig. 11 True Vs the predicted values of the reactive power by the hybrid transformer-TCN model

The more significant difference between the true values and the model is observed in the period of the peaks and the valleys, indicating that although the model is fairly accurate within the normal range, it may be somewhat less successful in incorporating extremes within the reactive power range. However, as depicted by most of the areas having a tight proximity of the lines, the Transformer-TCN hybrid model performed well in predicting reactive power, including both short and long temporal features. As this graph shows, the model can work with a fair amount of accuracy, but there is always room for reducing variability, especially in the outliers.

Since the Transformer-TCN model used for reactive power forecasting contains features such as Global Reactive Power, Global Active Power, Voltage, and Global Intensity, the heatmap of Figure 12 also shows the correlation between these features. The color scheme represents the strength and sign of the correlations: darker reds for positive values and blue tones for negative values.

One thing that is quite striking is the negative relationship between Global Active Power and Voltage, which is -0.63, meaning that as active power goes up, the voltage goes down. Likewise, both Global Intensity and Voltage negatively correlate with -0.64. On the other hand, the correlation between Global Active Power and Global Intensity is positive, equal to 1.0, which indicates direct dependency. As for the correlation coefficients, Global Reactive Powers has a moderate positive correlation with Global Intensity of 0.26 and Global Active Power of 0.25, implying that active power and intensity go up, and so does the reactive power to a certain extent.

The heatmap is suitable for revealing correlations between the features, which is important for understanding how these variables are related and how they impact the reactive power prediction in the Transformer-TCN model. While these relationships can help explain the model's ability to accurately capture patterns in the power system dynamics, there is no guarantee that they will do so.

In the graph of Figure 13, the residuals represent the number of true values and the predicted ones for the Transformer-TCN model for the initial 100 samples. The residuals, shown as the red line at the bottom of the plot, give more information about the model's prediction accuracy by showing the difference between the given and predicted reactive power.

In most of the samples, the residuals deviate only slightly from zero, which suggests that the model does a good job of predicting reactive power. Certain fluctuations exist; the increases and decreases are most significant in samples 20, 60, and 90. These spikes indicate the fact that the model is not very accurate in predicting reactive power at points where it experiences sharp fluctuations or is at its extreme value. The residuals remain generally small for most samples; however, large deviations in several specific cases point to avenues for model enhancement, especially regarding rapid variability in reactive power.

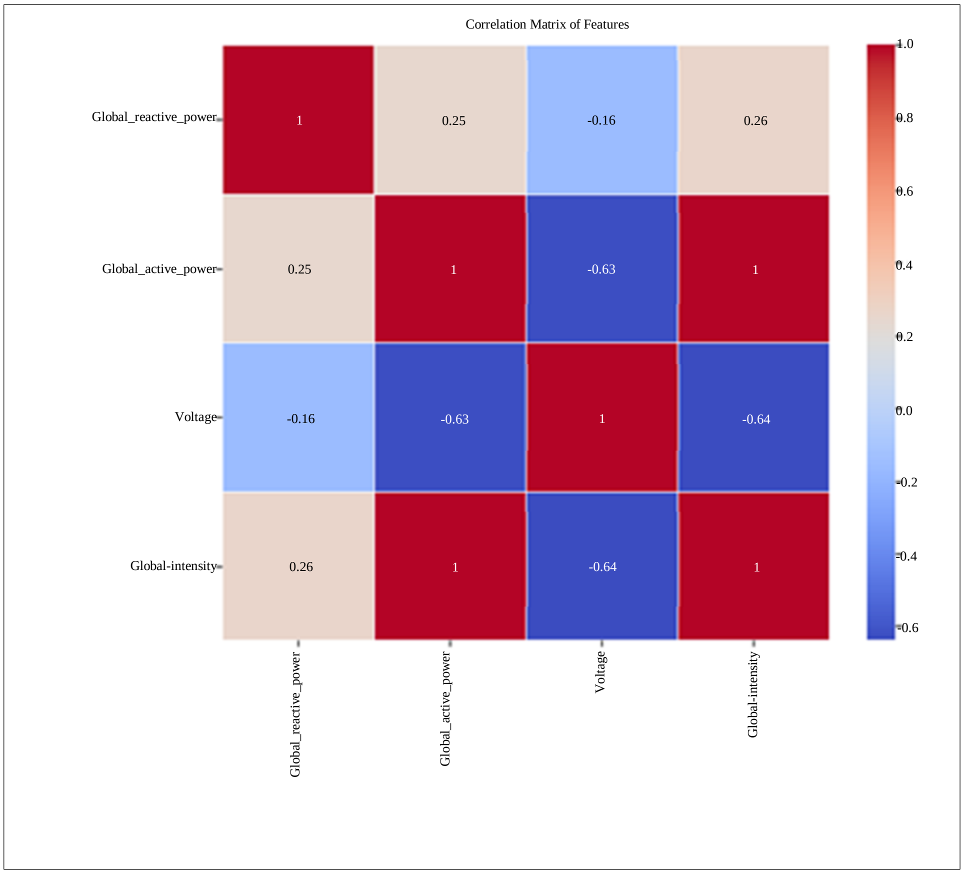

Fig. 12 The heatmap of the hybrid transformer-TCN model

The data distribution in Figure 14 represents the true and the predicted reactive power of the Transformer-TCN model on the x and y axes, respectively. In the ideal case where the model is perfectly accurate, the points should lie on the diagonal line where the predicted values are true.

In this plot, most points lie in the 0.1-0.4 kVAR range. Hence, the model is fairly uniform in predicting reactive power within this range. However, as the actual reactive power increases beyond 0.4 kVAR, the scatter of the points increases, which leads to the conclusion that the model cannot predict higher reactive power values. Furthermore, a concentration of data points in a vertical line at 0 kVAR on the x-axis suggests that the model at times yields estimates close to 0.3 kVAR for those occasions when the true reactive power is very low.

The scatter plot further reveals that, although the Transformer-TCN model meets expectations for moderate reactive power levels, it shows higher errors for extremes, especially as true reactive power increases. This implies that there is still great potential to refine the proposed model to provide a more accurate representation of reactive power as a function of voltage, particularly for values that fall outside the range encompassed by the data used in developing the model.

The histogram in Figure 15 demonstrates the error distribution (residuals) of the Transformer-TCN model that illustrates discrepancies between real and forecasted reactive power. The plot is symmetrical around the origin, which means we can affirm that, on average, the model's errors are correct, and they balance each other. However, the residue has positive skewness, which suggests that there are more cases where the model overestimates the actual reactive power than where it is underestimated.

Looking at the distribution of the residuals, it is seen that most of the values are between -0.1 and 0.1. Therefore, most of the predictions have small errors. This implies that for most of the data points, the model gives small values of prediction errors. However, some values greater than 0.4 go toward the model to predict some data type, which causes the model to fail on a few data points, making its prediction error higher than others.

The distribution gives a general idea of the model's performance: it can be seen that the Transformer-TCN model works quite well in general; however, there can be more significant errors, presumably in more specific and extreme cases. Getting a more precise parameterization of the model could also help decrease such large errors.

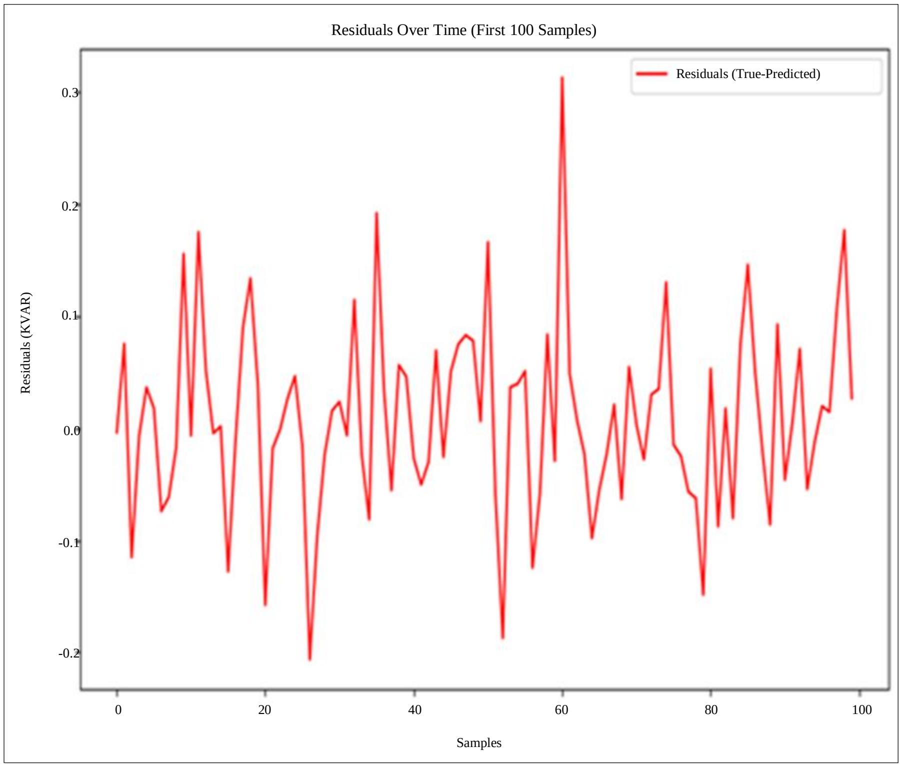

Fig.13 Residuals represent the true values and the predicted ones for the transformer-TCN model

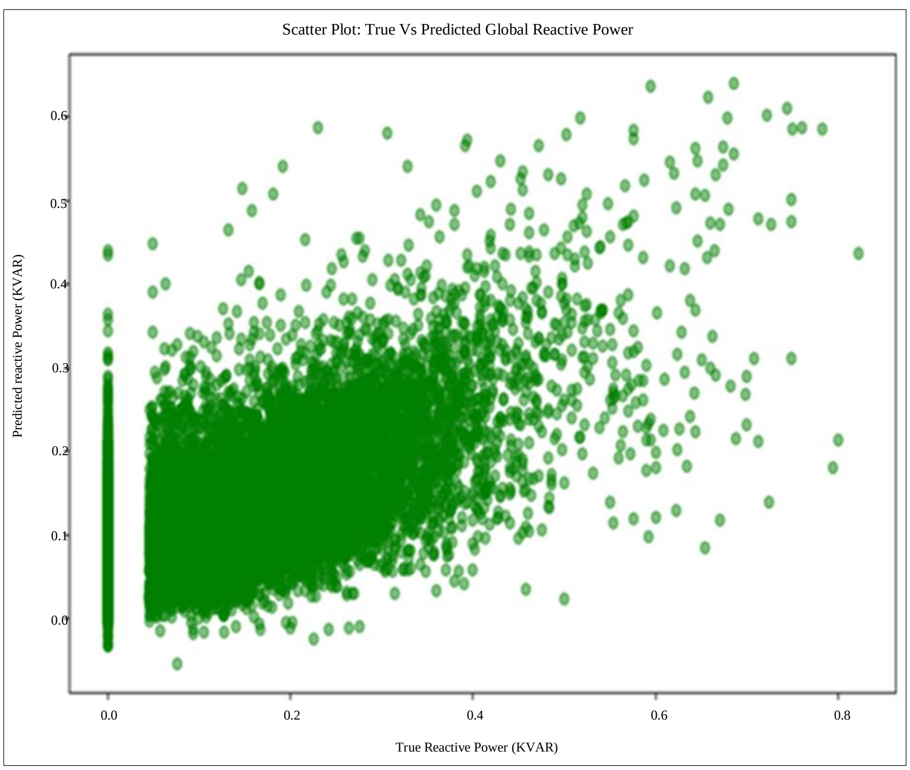

Fig. 14 Scatter plot of the true Vs the predicted values of the reactive power by the hybrid transformer-TCN model

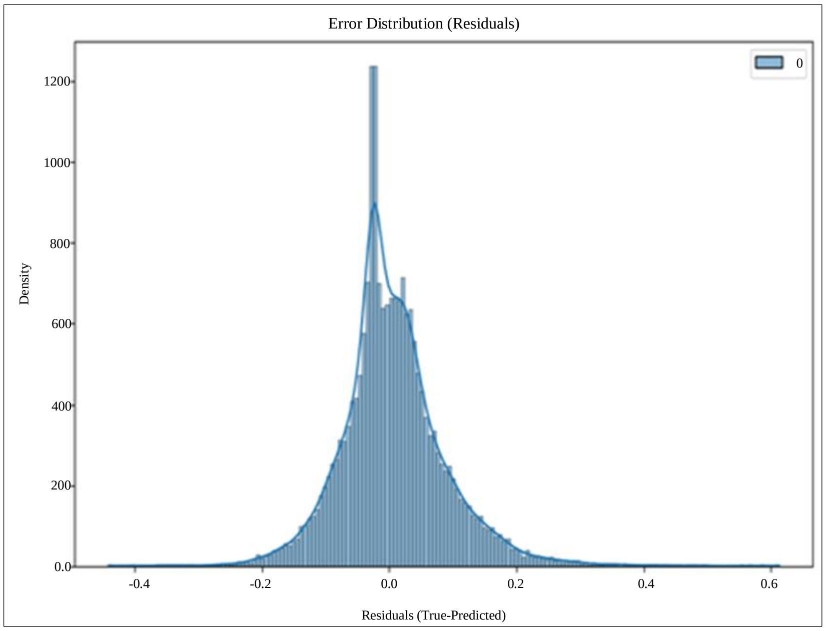

Fig. 15 Error distribution (residuals) of the transformer-TCN model

7. Conclusion

This paper analyzed the potential of several deep learning models, namely, TCN, Transformer, TCN-Transformer, and Transformer-TCN, in the case of reactive power forecasting in power systems. In light of these models, the performance of each model was compared and evaluated using metrics such as MAE, MSE, and RMSE. The Transformer-TCN combined model showed the highest performance across all the tested models' metrics. This model effectively leveraged the strengths of both architectures: the long-range dependency capturing ability of the Transformer at the global level and the short-term temporal features learning ability of TCN at the local level. What made the Transformer-TCN model better was that it provided more accurate and precise predictions of the reactive power, reducing the absolute and squared errors.

The analysis also showed that an average of the models offered better results than each of the models separately. As both TCN and Transformer models demonstrated promising results in capturing reactive power patterns, integrating the two architectures exhibited better performance because the time series learning in the power system is more complicated. The TCN-Transformer model was also accurate but was not as close to the reverse hybrid, which again proved that the sequence in which the models are merged is critical. More specifically, true vs. predicted value plots, residual plots, error distributions, and correlation matrices obtained from visual analysis supported the models' predictive abilities. It was ascertained that the Transformer-TCN model performed well in tracking the actual reactive power trends, but certain issues were identified with the model in predicting large deviations, as highlighted by the residual and scatter plots. Self-check analysis of the model showed that the residuals clustered around the zero line, implying the model's overall reliability. Therefore, the study emphasizes the benefits of combining deep learning models in the context of RPQ in today's emerging power system. In addition to enabling shorter and longer-term dependencies, a model of that kind demonstrated higher accuracy and stronger stability than separated architectures. These outcomes indicate that more future research could improve reactive power forecasting by investigating other configurations and hyperparameters or combining them with other superior methods. Further, the findings from this study may help refine demand forecasting techniques to enhance the power grid's function and maintain system stability.

Funding Statement

This study was performed with Jamhuriya University of Science and Technology, Mogadishu, Somalia.

Availability of Data and Materials

The data are available online as in reference [20].

Acknowledgments

The authors would like to thank the engineering faculty of Jamhuriya University of Science and Technology for their invaluable assistance and support throughout this research. Special thanks to the Department of Electrical Engineering for their guidance and contributions. The authors also appreciate the cooperation and encouragement from their colleagues and peers.

References

[1] Amir Sepehr, Oriol Gomis-Bellmunt, and Edris Pouresmaeill, "Employing Machine Learning for Enhancing Transient Stability of Power Synchronization Control During Fault Conditions in Weak Grids," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2121-2131, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Wenping Qin et al., "Reactive Power Aspects in Reliability Assessment of Power Systems," *IEEE Transactions Power System*, vol. 26, no. 1, pp. 85-92, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[3] Iván Andrade et al., "An Active/Reactive Power Control Strategy for Renewable," *Electronics*, vol. 10, no. 9, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Antonio Bracale et al., "Short-Term Industrial Reactive Power Forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 177-185, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jinling Lu et al., "Two-Tier Reactive Power and Voltage Control Strategy Based on Arma Renewable Power Forecasting Models," *Energies*, vol. 10, no. 10, pp. 1-13, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ahmad Rofii et al., "Utilize the Prediction Results from the Neural Network Gate Recurrent Unit (GRU) Model to Optimize Reactive Power Usage in High-Rise Buildings," *International Journal of Robotics & Control Systems*, vol. 4, no. 2, pp. 628-654, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Mohamed Massaoudi et al., "Fast Transient Stability Assessment of Power Systems Using Optimized Temporal Convolutional Networks," *IEEE Open Journal of Industry Applications*, vol. 5, pp. 267-282, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sami Ullah et al., "Development of a CNN-based Time-Varying Reactive Power Forecast Model for Power systems," *2022 International Conference on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS)*, Islamabad, Pakistan, pp. 1-7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Hatice Vildan Dudukcu et al., "Temporal Convolutional Networks with RNN Approach for Chaotic Time Series Prediction," *Applied Soft Computing*, vol. 133, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Huiling Chen et al., "Early Time Series Classification Using TCN-Transformer," *2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Dali, China, pp. 1079-1082, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xiaoyan Tian, Ye Jin, and Xianglong Tang, "TSRN: Two-Stage Refinement Network for Temporal Action Segmentation," *Pattern Analysis and Applications*, vol. 26, no. 3, pp. 1375-1393, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Olof Almqvist, "*A comparative study between algorithms for time series forecasting on customer prediction: An investigation into the performance of ARIMA, RNN, LSTM, TCN and HMM*, University of Skövde, School of Informatics, Student Thesis, 2019. [Google Scholar] [Publisher Link]

[13] Shilin Sun et al., "Short-Term Multi-Step Wind Power Forecasting Based on Spatio-Temporal Correlations and Transformer Neural Networks," *Energy Conversion and Management*, vol. 283, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv preprint*, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] André Veltman, Duco W. J. Pulle, and R. W. De Doncker, *The Transformer*, Power Systems, Springer, Cham, pp. 47-82, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] Guokun Lai et al., "Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks," *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, New York, United States, pp. 95-104, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[17] Gülizar Gizem Tolun, and Kasım Zor, "Very Short-Term Reactive Power Forecasting Using Machine Learning-Based Algorithms," *2024 9th International Youth Conference on Energy (IYCE)*, Colmar, France, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Renzhuo Wan et al., "Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting," *Electronics*, vol. 8, no. 8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Heng Yuan et al., "Vehicle Trajectory Prediction Based on Posterior Distributions Fitting and TCN-Transformer," *IEEE Transactions on Transportation Electrification*, vol. 10, no. 3, pp. 7160-7173, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Georges Hebrail, and Alice Berard, "Individual Household Electric Power Consumption," *UCI Machine Learning Repository*, 2006. [CrossRef] [Google Scholar] [Publisher Link]