IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 July 2024, accepted 5 August 2024, date of publication 7 August 2024, date of current version 16 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3440226

RESEARCH ARTICLE

Adaptive Renewable Energy Forecasting Utilizing a Data-Driven PCA–Transformer Architecture

FAHMAN SAEED AND SULTAN ALDERA

Computer Science Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

Corresponding author: Fahman Saeed (faesaeed@imamu.edu.sa)

This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) under Grant IMSIU-RG23094.

ABSTRACT The incorporation of renewable energy sources into the power grid has necessitated the development of sophisticated forecasting models that can effectively handle the inherent fluctuation and uncertainty associated with renewable energy generation. In this study, an adaptive principal component analysis (PCA)-enhanced transformer architecture, hereinafter referred to as PCA–Transformer, is developed to enhance the performance of transformer models in predicting renewable energy output. The proposed model uses PCA to dynamically determine and adapt the transformer architecture and prioritize the most informative features from time series data, thereby improving the model's attention on relevant information and reducing computational burden. This is essential for accurately capturing the intricate temporal patterns and nonlinear relationships that are typical present in renewable energy time series data. The PCA–Transformer enhances the performance of transformer models in sequence-to-sequence predictions by incorporating an adaptive mechanism that customizes their structure based on the best PCA eigenvectors. The architecture adaptively aligns with the underlying patterns in data by adjusting the number of attention heads and critical dimensions within each transformer block. The adaptability of the proposed architecture is crucial for effectively simulating the complex nature of renewable energy generation patterns. The efficiency of the proposed model was evaluated using the Alice Springs Australia DKASC-ASA and EIA Energy datasets. The proposed model has superior forecasting performance than traditional transformer models and cutting-edge renewable energy forecasting methodologies.

INDEX TERMS AutoML, principal component analysis, renewable energy forecasting, time series analysis, transformer.

I. INTRODUCTION

The increasing environmental concerns and global need for sustainable energy solutions has necessitated the use of renewable energy sources. Power generation has increased due to energy dependency [1]. Fossil fuels cause global warming, air pollution, and climate change by emitting greenhouse gases [2]. With the promotion of renewable energy technologies and the near-zero carbon emission policy, green development has become the cornerstone of livelihood. Therefore, many countries have focused on using renewable energy systems (RESs) to generate electricity from renewable sources [3]. The International Energy Agency predicts that ~60% of electricity generated in 2040 will from renewable sources [4]. Renewable energy, which includes solar, wind, and hydroelectric power, can reduce greenhouse gas emissions, prevent climate change, and promote socio-economic well-being. However, the intermittency of renewable energy sources, which is closely linked to factors such as weather patterns and environmental conditions, significantly limits their extensive implementation [5]. Such intermittencies can cause energy shortages, grid instability, and suboptimal utilization of these eco-friendly resources. Thus, the optimization and forecasting of renewable energy generation is a pivotal obstacle hindering the realization of a greener and more sustainable energy ecosystem. With

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali.

increased use of renewable energy sources, effectively managing and forecasting their output is crucial. When renewable energy generation is reliably predicted, grid operators can balance supply and demand efficiently, optimize energy storage systems, and enhance grid stability. However, obtaining accurate and dependable renewable energy forecasts in the presence of dynamic and interrelated factors is a formidable task.

Therefore, several studies have proposed methods to improve the precision of forecasting systems. For instance, the reliability of forecasting methods can be enhanced by correlating temporal lags and observations [6]. Time series data from solar or wind energy systems exhibit various patterns but also some similarities. The systems may have numerous connected time series or exogenous factors that improve the forecasting model [7], [8], [9]. There exist uncertainties in the time series data caused by climate or power system variations, which can influence a model designed to handle a specific pattern [10]. Dimensionality reduction in time series forecasting is another common problem, which can be efficiently solved using principal component analysis (PCA) [11]. PCA simplifies complex models by focusing on the most important variables and helps attenuate noise, yielding more accurate forecasts [12]. It can also isolate essential predictive traits, which is a valuable ability in the analysis of time series [13]. Owing to its high computational efficiency, PCA can be applied on large datasets [14] and addresses collinearity in multivariate time series [15]. The applicability of PCA goes far beyond the realms of economics, energy, and finance; it also finds use in environmental science, where it facilitates the modeling and forecasting of climate [16]. The role of PCA has advanced due to the emergence of machine learning (ML); PCA is now used as a pre-processing step for enhancing advanced algorithms such as support vector machines (SVMs) and neural networks, enabling the use of more sophisticated and precise forecasting approaches [17].

Transformers have significantly advanced the practice of forecasting across domains such as finance, energy, weather, and sales because they can effectively handle lengthy sequences, analyze data simultaneously, and provide enhanced precision compared to conventional models such as ARIMA [18] or LSTM [19]. They can capture intricate long-range dependencies owing to their superior attention mechanisms, efficiently handling complex time series data [20]. They can also expand and adapt to meet different forecasting task requirements and incorporate more amount of data for a more complete image. However, model training may involve overfitting because large quantities of data are used, which increases the computational cost [21]. Nevertheless, transformers are a noteworthy advancement in time series forecasting, for which transformer models must be carefully architecturally modified to accommodate sequential data. Wen et al. [22] reported that module-level adaptations allow intricate modifications via inductive biases suitable for temporal data. Moreover, the processing of time series input can be improved using novel attention mechanisms and normalization approaches. The architectural-level models are inherently efficient and scalable and maintain linear computational complexity while processing huge amounts of time series data. Seasonal-trend decomposition in the transformer framework improves forecasting accuracy by 50%–80% [23]. The length of the input sequence is also important; as longer sequences enable the model to effectively capture long-range dependencies. The proportions of the model also exert a substantial influence. Shiyang et al. [24] examined the relation between the model size and forecasting efficiency and reported that transformers with a lower number of layers (3–6) exhibit superior forecasting performance than those with a higher number of layers. Haixu et al. [25] incorporated periodicity and trends directly into the transformer architecture to enhance its forecasting performance. Transformers and graph neural networks are effective for spatio-temporal forecasting, and transformer models have been extensively studied recently for predicting long-term time series. The LogTrans model, introduced by Li et al. in 2019 [24], uses convolutional self-attention layers with a LogSparse architecture to effectively capture local information and reduce spatial complexity. The ProbSparse self-attention mechanism uses distillation techniques to effectively extract the most important keys [26]. Wu et al.'s Autoformer model [25] integrates the principles of decomposition and auto-correlation derived from traditional time series analysis approaches. The FEDformer model [27], proposed by Zhou et al. in 2022, uses a Fourier-enhanced structure to achieve linear complexity. Liu et al.'s Pyraformer [28] uses a pyramidal attention module that incorporates inter- and intra-scale connections, resulting in linear computational complexity. Automated ML (AutoML) automates every ML step, from data preparation and feature engineering to model generation and evaluation [29], [30]. AutoML can also be used to automatically construct the architectural layers and filters of a model based on the provided data. It considerably improves the utilization of neural network models such as transformers in several tasks, including forecasting, by automating intricate operations such as model selection, optimization, and tweaking, thereby improving the accessibility of the transformers' superior capabilities to a broader user base as well as their efficiency and performance. AutoML is particularly useful in forecasting, wherein transformers perform exceptionally well because they can handle long sequences and intricate relationships. It simplifies the complexity of these models; thus, they can be deployed in various scenarios without the requirement of extensive technical knowledge. The incorporation of AutoML with transformers signifies a pivotal advancement in enhancing the accessibility and applicability of sophisticated forecasting methodologies across several fields.

Pre-existing AutoML technologies [31], [32] can be used for enhancing the forecasting performance as they accelerate the search for the best model architecture for classification problems using DNNs with different types of layers. AutoML libraries can deploy many AutoML algorithms on the target data and perform tasks, including automated search for the best model architecture, identifying model ensemble, and performing model distillation. Regularized evolution can be used to modify neural network architecture. As each gene encodes the number of filters in a layer, an individual can be characterized by a three-layer neural network sequence of integers. To develop a customized AutoML system using the transformer architecture [30], the model must be programmed and techniques such as positional encoding must be used for the model to predict the constituent order of sequences.

The advantages of PCA and transformer models can be combined to obtain a data-driven automated transformer. PCA is particularly advantageous in dealing with the complexity of large RESs because it can successfully reduce the number of dimensions and identify important characteristics from large datasets. Transformer models achieve accurate forecasting as they can adeptly process sequential data, identify complicated patterns in time series data, and capture long-term relationships. Such integrated models yield more accurate and reliable predictions than PCA and transformers used alone. They are highly adaptable and can dynamically modify their design and parameters based on the input data, enabling their continual evolution and improvement in handling dynamic renewable energy data. This ability is essential in an industry marked by swift fluctuations in circumstances and consumer behavior. Moreover, data-driven automated transformers can scale and operate efficiently; therefore, they are highly suitable for handling expansive datasets common to renewable energy applications. This ensures its effectiveness and suitability within a progressive framework.

The PCA-Transformer model creatively incorporates Principal Component Analysis (PCA) to dynamically adjust the structure of transformer models in order to improve the accuracy of predicting renewable energy outputs. This model utilizes Principal Component Analysis (PCA) to adaptively modify its configuration, specifically the number and dimensions of attention heads, based on the importance of features detected in the data. This modification enables the transformer to systematically improve its structure, one layer at a time, resulting in a notable increase in efficiency by minimizing unnecessary calculations and allocating processing capacity to the most influential characteristics. Furthermore, this model has exceptional performance in predicting accuracy by precisely adjusting each layer to efficiently handle diverse data complexities and scales. The PCA-Transformer's capacity to adjust to new data as it becomes accessible guarantees its resilience and significance in dynamic settings such as renewable energy, where data attributes often change. This approach enhances both the operational efficiency and forecasting accuracy of the model, while also providing notable flexibility, representing a significant improvement over conventional static transformer models. The adaptive nature of this approach guarantees the ability to handle large amounts of data and adapt to different situations, making the PCA-Transformer especially suitable for the intricate task of modeling and forecasting trends in renewable energy data.

Our method synergistically integrates PCA and Transformer technology, leveraging PCA to dynamically adapt the Transformer architecture for the specific purpose of renewable energy forecasting. This approach tackles the substantial difficulties associated with forecasting renewable energy outputs in the face of intermittent oscillations that can lead to energy deficits, grid instability, and suboptimal utilization of renewable resources. These oscillations present significant obstacles to establishing a sustainable energy ecology. The PCA-Transformer model improves the accuracy and dependability of these forecasts, which is essential for optimizing grid operations. The system utilizes Principal Component Analysis (PCA) to discern and prioritize crucial characteristics from time-series data. This enables grid operators to effectively handle the balance between supply and demand, maximize energy storage, and uphold grid stability, even in the face of the fluctuating nature of renewable energy sources.

The interpretability of our PCA-Transformer model is improved by utilizing PCA to identify and prioritize important features prior to their processing by the Transformer architecture. This approach enables us to clearly identify the features that have a substantial impact on the forecasting results, hence improving the transparency of the model's decision-making process. During the Transformer stage, the attention processes are influenced by these prioritized features, offering valuable insights into the contribution of certain data items to the final predictions. By incorporating Principal Component Analysis (PCA) directly into the design process of the Transformer architecture, we may dynamically modify the architecture during training. Conventional approaches frequently depend on manual experimentation or rule-based adjustments to enhance model structure, such as modifying the number of attention heads and layer weights. However, these methods can be time-consuming and imprecise. Our approach use Principal Component Analysis (PCA) to dynamically adjust the Transformer design by consistently finding the most crucial components of the input. It dynamically adapts the number and sizes of attention heads to ensure that each layer is properly tuned to concentrate on the most crucial aspects, while minimizing computational redundancy by giving fewer resources to less critical data points. This not only enhances the efficiency and accuracy of the model when dealing with fluctuating renewable energy data, but also improves the interpretability of the model and helps achieve broader sustainability goals by optimizing energy forecasting processes. This adaptable strategy not only tackles the intricacies of forecasting renewable energy in the midst of variable and interconnected circumstances, but also supports sustainability goals by enhancing the economic and environmental results of using renewable energy. Therefore, our model provides a methodical answer to the problem of architectural determination in AutoML, delivering a versatile, data-driven, and significantly more efficient approach.

Our research contributions involving the development of a data-driven model for renewable energy forecasting using a data-driven automated transformer are as follows:

1. **Dynamic Optimization of Model Architecture:** We have created a customized forecasting model that adapts its structure by incorporating Principal Component Analysis (PCA). Applying PCA to the training data allows our model to determine the appropriate degree of complexity required for accurate forecasting. Additionally, PCA helps to accurately adjust the weights across each attention layer. This customized technique guarantees that the model's structure is tailored to effectively capture and highlight the most significant characteristics of the data.
2. **Enhanced Forecasting Accuracy and Reliability:** The PCA-Transformer model involves thorough training using raw time series data and demanding validation using tested datasets to guarantee its performance. This strategy greatly enhances both the accuracy and reliability of predicting renewable energy production. Our approach improves energy management and planning by accurately predicting outcomes through the analysis of the key variables that contribute to the majority of variance in the data.

These distinct PCA integrations into transformer architectures increase forecasting precision and model adaptability, which are essential for renewable energy consumption and sustainable energy systems. The model's rapid processing of huge datasets and adaptation to changeable energy data make it scalable and applicable in sustainable energy applications. Thus, it aids sustainable energy management, particularly renewable energy integration and use.

This paper is organized as follows: **Section I** contains the introduction, **Section II** contains the proposed method, **Section III** contains the experimental results, and **Section IV** contains the conclusion.

II. PROPOSED METHOD

Data-driven time series forecasting aids researchers in understanding system evolution without using physical laws. Time series models have undergone several developments to meet diverse industry requirements. However, constructing transformer neural architectures are difficult because the best network depth, embedding dimension, and head number have to be determined. These parameters are crucial for increasing the model capacity; however, achieving their suitable combination is challenging. Transformers with higher depths, head numbers, and multilayer perceptron (MLP) ratios (the ratio of hidden dimension to embedding dimension) offer improved prediction accuracy but overfit after peaking. The model performance can be improved by scaling up the embedding dimension; however, the accuracy plateaus for larger models. These phenomena demonstrate the complexity of a transformer architecture design. Previous transformer designs were built manually by experienced personnel using trial-and-error approaches. Herein, two transformer search challenges are addressed: how to (1) balance essential transformer factors such as net depth, embedding dimension, and head number and 2) customize transformer models to be efficiently identified for the target renewable energy dataset. Substantial efforts have been put to optimize the transformers models such as LogTrans [24], Autoformer [25], Triformer [33], FEDformer [27], and Pyraformer [28]. LogTrans [24] eliminates the need for a point-wise dot product between the key and query, but its output is determined by a single time step. Autoformer [25] employs auto-correlation to establish connections at the patch level; however, it is a manually designed model that does not encompass all the semantic information contained within a patch. Moreover, Triformer [33] uses patch attention to decrease complexity but employs a pseudo timestamp as the query within a patch, i.e., it does not consider a patch as an input unit or provide insight into its semantic significance. Herein, an advanced data-driven automated transformer is proposed to estimate the production of renewable energy. The optimized model design uses PCA to establish the ideal model complexity level. PCA identifies the most important training data properties, forming a customized forecasting framework. The proposed model is trained using original data and validated using tested datasets to improve its prediction accuracy of renewable energy generation as well as learning and validation. The model can efficiently analyze large datasets and react to varying energy data. Thus, it can improve renewable energy management, integration, and use, highlighting its potential, scalability, and versatility in sustainable energy projects. The original transformer block used herein is the one introduced in [34]. The dynamic transformer architecture for time series forecasting of renewable energy is shown in Algorithm 1, and the entire process is schematized in Figure. 1. Step 1 involves initializing the initial number of heads (initial_num_heads) and key dimension (initial_key_dim) as well as hyper parameters. The hyper parameters play a critical role in determining the model's capacity to acquire knowledge from the training data. The variable "best_loss" is set to an infinitely large value for early halting to mitigate overfitting. An input tensor is then generated, which has the same dimensions as the training data (X_train); this process serves as the input layer for the model. After establishing the framework, the model creation process commences. No transformer blocks are available when the model creation process begins, thus providing a fresh starting point for the iterative development process. Then, additional transformer blocks are incorporated into the model architecture by successively increasing the model depth using B transformer blocks one at a time in a loop. In In Step 2, PCA is applied to input training data after the initial format if needed, and PCA is used to extract the principal iteration. The previous_block_output is reshaped into a 2D components, also known as eigenvectors. This procedure is crucial as the most important characteristics are extracted from the output of the previous block, which influences the design of the next block. To obtain the first set of eigenvectors, PCA is applied to the reorganized X_train

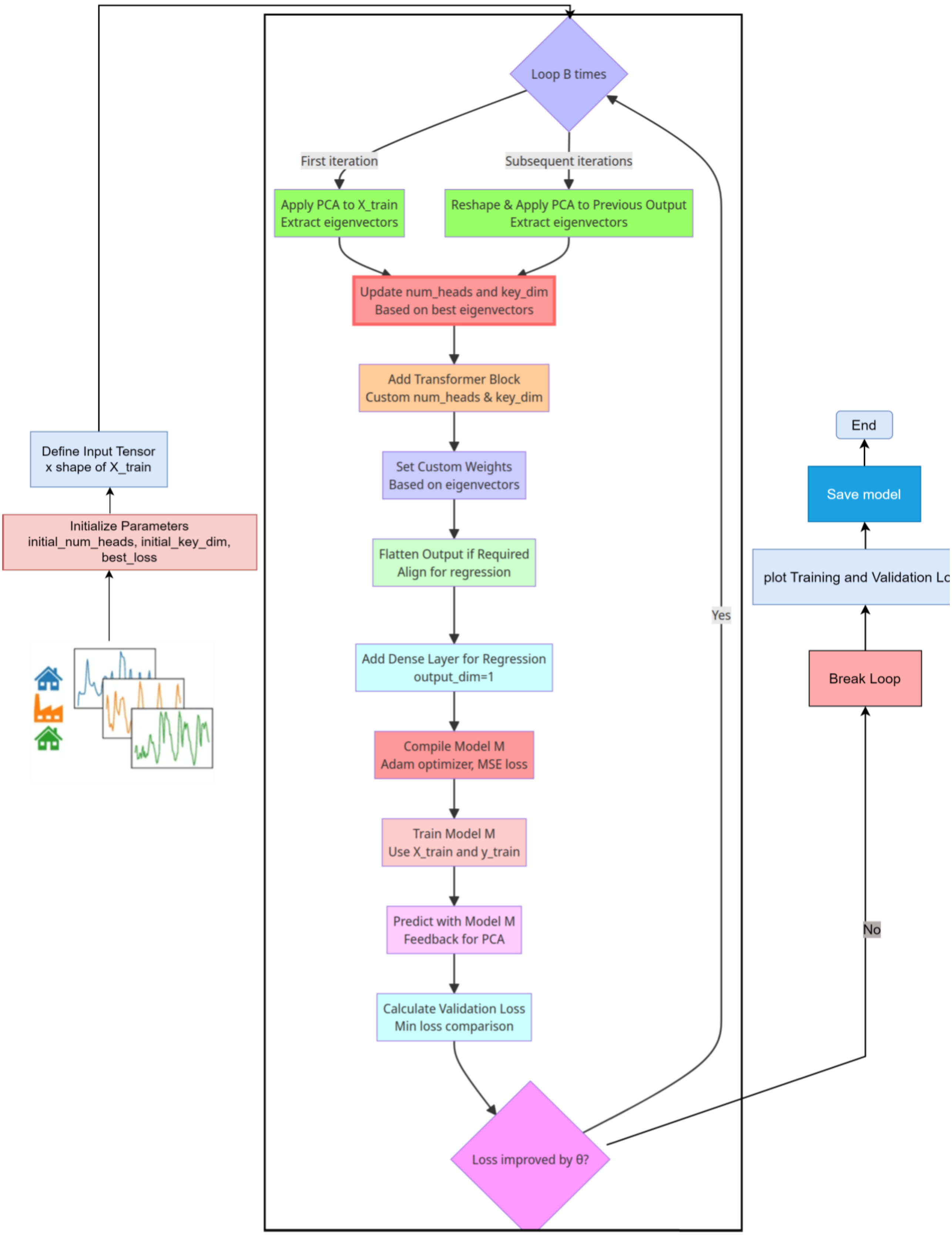


FIGURE 1. PCA–transformer architecture.

Algorithm 1 Dynamic Transformer Architecture for Time Series Forecasting

Input: Training dataset X_train, corresponding target values y_train
Improvement threshold θ, desired number of blocks B
Output: Trained forecasting model M

1: Initialize initial_num_heads, initial_key_dim
Initialize best_loss to ∞
Define input tensor with shape of X_train
Let x be the input tensor
Let previous_block_output be None
for i = 1 to B do
2: if i > 1, then
Reshape previous_block_output to 2D if necessary
eigenvectors ← PCA(previous_block_output)
else
eigenvectors ← PCA(X_train reshaped to 2D)
end if
3: Update current_num_heads and current_key_dim based on the number of best eigenvectors which computed using:
a. we calculate the explained variance for each component as the ratio of its eigenvalue λ_i to the total sum of all eigenvalues $EV_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$
b. we retain the smallest number k of components such that the cumulative explained variance $CEV_k = \sum_{j=1}^{k} EV_i$ exceeds a threshold (95%).
4: x ← Addtransformer Block(x, current_num_heads, current_key_dim)
5: if i >= 1 then
SetCustomWeights(x, eigenvectors)
end if
6: if FlattenOutputRequired, then
x ← Flatten(x)
end if
7: outputs ← AddDenseLayer(x, output_dim = 1) for regression
8: M ← CompileModel(inputs, outputs)
9: Train model M using X_train and y_train
10: previous_block_output ← M.Predict(X_train)
11: current_loss ← MinValidationLoss(M)
if best_loss - current_loss > θ then
best_loss ← current_loss
else
12: break
end if
13: PlotTrainingAndValidationLoss()
end for
14: SaveModel(M)
return model M

dataset in Step 2 and the data are used to construct a baseline for the model. In Step 3, the block architecture is dynamically modified by updating the number of attention heads (num_heads) and the key dimension (key_dim) depending on the number of the best eigenvectors; to determine the best number of principal components, we calculate the explained variance for each component as the ratio of its eigenvalue λ_i to the total sum of all eigenvalues $EV_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$ and retain the smallest number k of components such that the cumulative explained variance $CEVk = \sum_{j=1}^{k} EV_i$ exceeds a threshold (95%) witch is 32 for encoder layer 1 of PCA–Transformer architecture for the DKASC-ASA-2 dataset (Figure. 3). This assures that retained components capture most of the data's volatility, reducing dimensionality while keeping important information.

In Step 4, the model incorporates a transformer block using appropriate values of num_heads and key_dim determined based on the number of the most optimal eigenvectors those preserve the maximum energy [35], [36]. In Step 5, to enhance the model's capacity to predict renewable energy

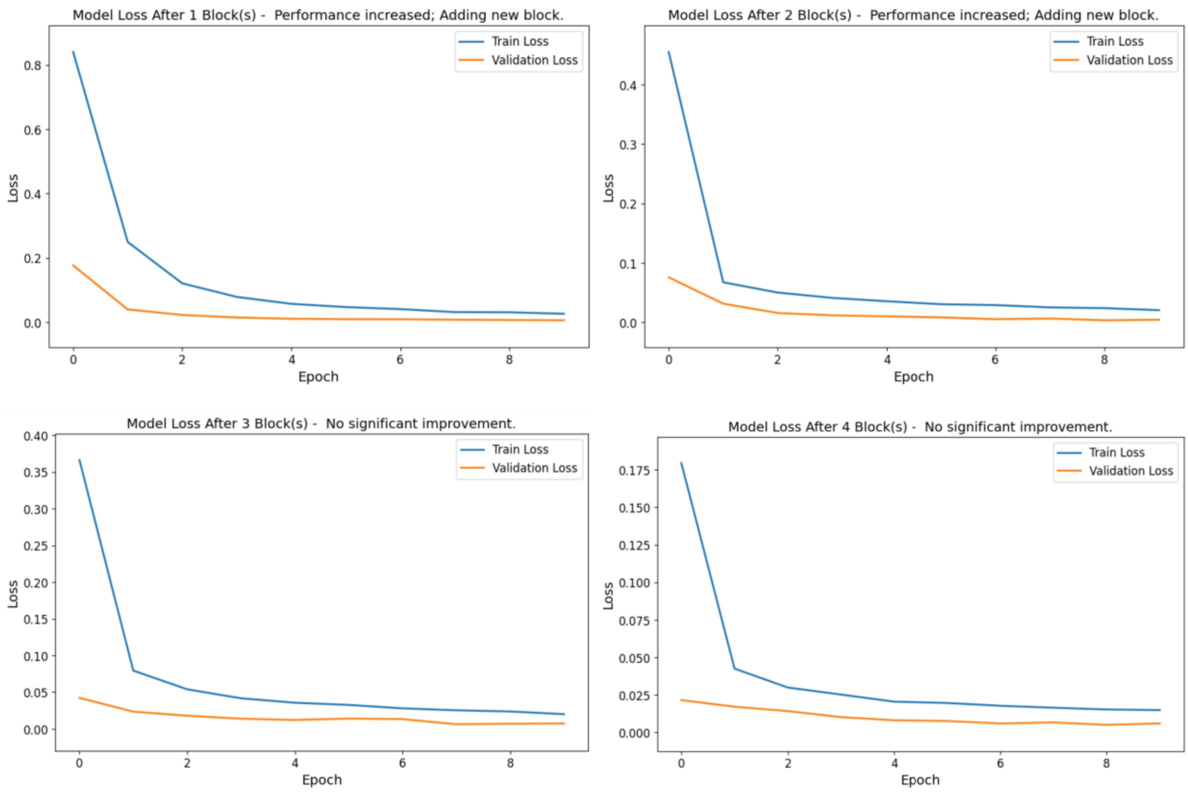


FIGURE 2. Stopping criterion of adding more layers in the PCA–Transformer model for the DKASC-ASA-2 dataset.

generation using representative features, the attention layers are configured with custom weights derived from the eigenvectors. This guarantees that every attention mechanism is in line with the most crucial aspects of the data, as identified by PCA (Steps 4 and 5 of the algorithm). Iteratively recalculating the weights at each epoch to accommodate for variations in data dynamics guarantees the model's flexibility and responsiveness to fresh data (Step 10 in the algorithm). This approach maximizes the intricacy and precision of the learning process by strengthening the model's capacity to concentrate on pertinent information and ensuring that it can adjust to evolving patterns over time. In Step 6, the output is compressed if needed to align with the regression framework, followed by Step 7, wherein a dense layer is introduced to combine the acquired features to obtain a predicted value. In Step 8, a model (M) is constructed using the Adam optimizer and mean-squared-error loss function because they are highly effective in regression tasks. In Step 9, M is trained to reduce the forecasting error, followed by Step 10, wherein M predicts the training data and creates a feedback loop critical for the subsequent PCA iterations. Step 11 involves determining whether another block needs to be added, which involves comparing the current validation loss with the best_loss at an improvement threshold θ (Figure. 2). An early stopping mechanism is implemented in Step 12 to prevent the model from becoming overly complex without any improvement in its performance. In Step 13, the training and validation losses are graphically plotted to visually track the model's progress and undertake necessary modifications. The iteration ends at Step 14 by preserving the trained model M, thereby ensuring that the acquired information may be accessed for future application or evaluation. This approach guarantees that the level of model complexity is justified by measurable enhancements in prediction accuracy, achieving a harmonious equilibrium between model sophistication and computing efficiency.

III. RESULTS AND DISCUSSION

Herein, the results of customizing PCA–Transformer using a target dataset are reported, and its performance is compared with those of other state-of-the-art transformer architectures used for time series forecasting.

A. DATASETS AND ADAPTIVE ARCHITECTURES

The effectiveness of the PCA–Transformer model was validated using three datasets: DKASC Alice Springs Australia DKASC-ASA-2, DKASC-ASA-1B [37], and EIA Energy dataset [38]; 80% and 20% of the data were used for training and testing, respectively. The DKASC data from ASA region hosts several operational solar power plants with varying generation capacities. The datasets contain data with a 5-min resolution. DKASC-ASA-2, with a generation capacity of 26.5 kW, was installed on August 23, 2010, whereas DKASC-ASA-1B, with a generation capacity of 23.4 kW, was installed on January 8, 2009. The data were collected since their installation until October 31, 2021. The dataset accessible through the U.S. Energy Information Administration (EIA) Total Energy data browser offers detailed information regarding the production and use of renewable energy categorized by source. The dataset spans from September 1985 to March 2024. The dataset offers extensive monthly data on several renewable energy sources, such as hydroelectric, wind, solar, geothermal, and biofuels. We will assess the performance of our model using data on wind and hydroelectric energy from this dataset. The performance of the PCA–Transformer model was assessed using commonly employed methods used for model evaluation: mean absolute error (MAE) method, which calculates the average of the absolute differences between the test set and the model output values, root mean square error (RMSE) method, which measures the square root of the average squared differences, and Coefficient of Determination (R^2) which indicates that the regression predictions perfectly fit the data. It assesses model accuracy in predicting future outcomes. The proportion of variation explained by the model indicates goodness of fit and how well unseen samples are likely to be predicted by the model.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}, \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}. \quad (2)$$

$$R^2 = 1 - \left(\frac{SS_{res}}{SS_{tot}}\right) \quad (3)$$

where n is the number of observations and y_i is the actual value and $\hat{y}_i$ is the predicted value. SS_res is the sum of squares of residuals (sum of squared differences between predicted and actual values) and SS_tot is the total sum of squares (variability of the data around the mean).

Using the Optuna optimization algorithm [39], We carefully adjusted the hyperparameters of our PCA-Transformer model in order to maximize its performance in predicting future outcomes. The optimization procedure entailed conducting experiments with three distinct optimizers—Adam, AdaFactor, and AdamW—across a range of learning rates spanning from 1e-5 to 1e-1. In addition, we assessed other activation functions such as ReLU, Leaky ReLU, and GELU, together with dropout rates ranging from 0.15 to 0.50. By conducting a thorough tuning process spanning 10 training epochs, we were able to determine the optimal combination of hyperparameters for each dataset. This meticulous adjustment of our model's architecture significantly improved its performance in predicting renewable energy. The optimized settings are outlined in Table 1, showcasing our detailed process in attaining the optimal model setup.

The PCA–Transformer model was data driven and customized for each dataset. The architecture obtained for the DKASC-ASA-2 dataset using Algorithm 1 comprised four attention layers (Figure.3(a)), whereas that for the DKASC-ASA-1B dataset comprised seven attention layers

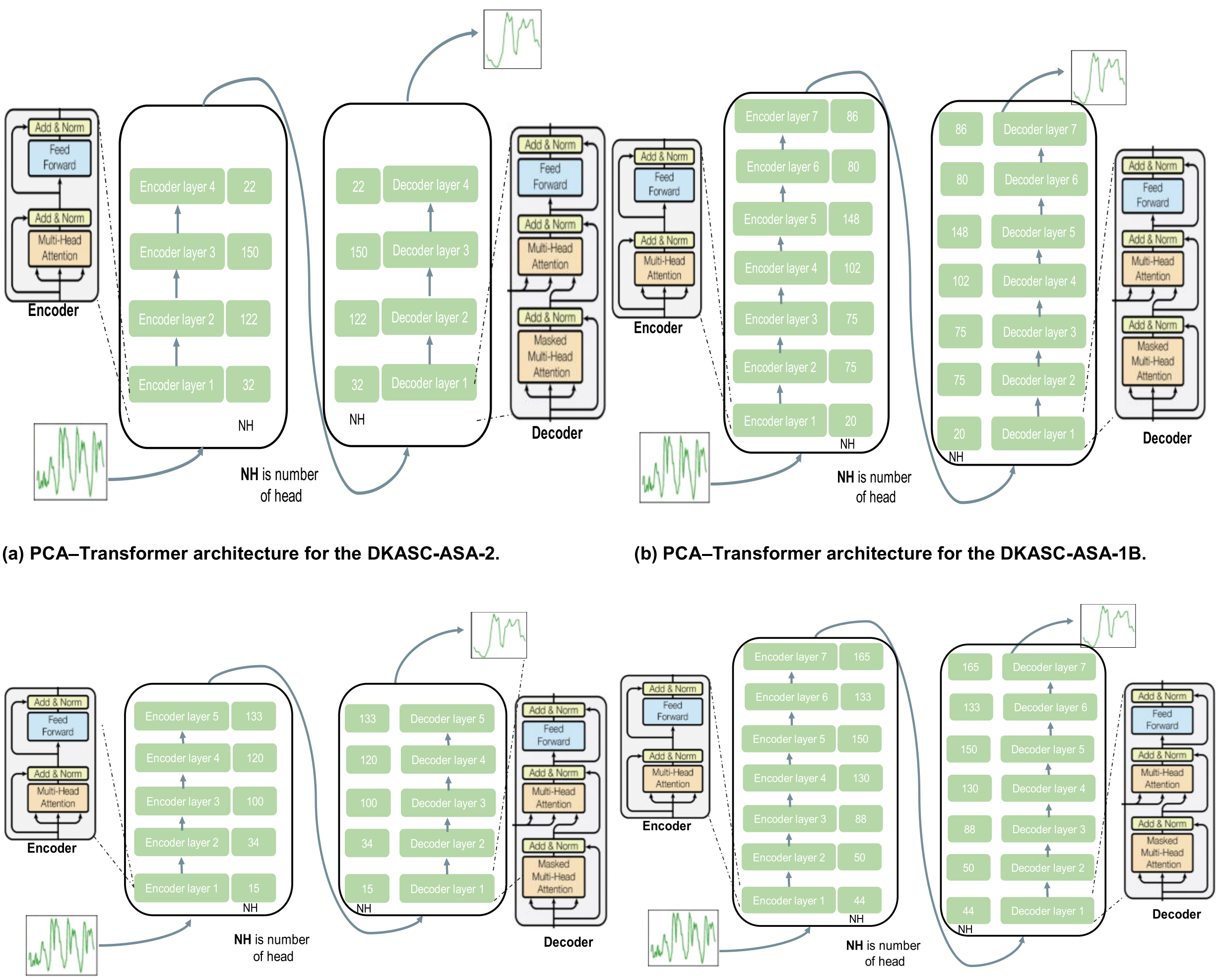
TABLE 1. Optimized hyperparameters obtained using the optuna optimization algorithm.

Dataset	Activation function	Learning rate	Optimizer	Dropout
DKASC-ASA-2	GELU	0.002	AdamW	0.25
DKASC-ASA-1B	GELU	0.007	AdamW	0.40
EIA wind energy	LRelu	0.006	AdamW	0.30
EIA Hydroelectric Energy	LRelu	0.006	AdamW	0.30

(Figure. 3(b)). The PCA–Transformer comprised five attention layers for EIA wind energy (Figure.3(c)), and seven attention layers for EIA Hydroelectric Energy (Figure.3(d)). To control the depth of the architecture, in Step 11 of Algorithm 1, the validation loss is compared with the best loss with an improvement threshold θ to determine whether another transformer block is required to be added. In steps 12, 13, and 14, an early stopping mechanism is implemented to stop the model from becoming overly complex without improving its performance and monitoring graphs of training and validation losses are obtained to track model progress as for DKASC-ASA-2 dataset (Figure. 2). To prevent the trained model from overfitting, the iterative process ends at depth 4 after the monitoring loss increases by three times. Thus, the suitable depth for the optimized PCA–Transformer model for the DKASC-ASA-2 dataset is determined as 4 layers. The number of attention heads for each layer and depth of each model for each dataset was adaptively determined based on the renewable energy dataset using Algorithm 1. The selection of the number of attention heads (num_heads) and the key dimension (key_dim) for each attention layer was dependent on the number of the optimal eigenvectors (Algorithm 1, Step 3). Owing to the distinct characteristics of the time series data of renewable energy sources, the number of layers and heads in the optimized model varied. To optimize the forecasting accuracy, a unique model configuration must be employed for each energy source category, such as solar or wind, and geographical area in the two datasets (Figure. 3(a: d)). These optimizations enhance the model's ability to handle inherent unpredictability and variability in the renewable energy data.

B. EXPERIMENTAL RESULTS

Herein, the empirical findings of the PCA–Transformer model developed for the two datasets are discussed. The proposed model outperformed other forecasting models when tested on the DKASC-ASA-2, DKASC-ASA-1B, EIA wind energy, and EIA Hydroelectric Energy datasets. Table 2 shows that the proposed model exhibited the highest accuracy and lowest error margins, with the best MAE, RMSE, and R^2 values being 0.0074, 0.11, and 0.98 respectively, for the DKASC-ASA-2 dataset. This implied that the model could identify and use data patterns. Autoformer



(a) PCA–Transformer architecture for the DKASC-ASA-2.

(b) PCA–Transformer architecture for the DKASC-ASA-1B.

(c) PCA–Transformer architecture for the EIA wind energy

(d) PCA–Transformer architecture for the EIA Hydroelectric Energy.

FIGURE 3. PCA–Transformer architecture of the different datasets (a): DKASC-ASA-2, (b): DKASC-ASA-1B, (c): EIA wind energy, and (d): EIA Hydroelectric Energy.

exhibited the second-best performance, whereas FEDformer and Pyraformer exhibited slightly higher error metrics; however, the former slightly outperformed the latter. The PCA–Transformer model outperformed the other models with large leads on the DKASC-ASA-1B dataset, with MAE, RMSE, and R^2 values being 0.12 and 0.2, and 0.95, respectively. Among all models, ARIMA performs the worst. The higher MAE and RMSE values for Pyraformer, FEDformer, and Autoformer indicate either a mismatch between the models' capabilities and the datasets' attributes or that the datasets were complex. The PCA–Transformer model used PCA to capture crucial elements and employed them as weights for the transformer's layers; it adaptively set the model depth that enabled more exact predictions and possibly higher generalization from the training data, indicative of its extraordinary performance. Autoformer performed well on the DKASC-ASA-2 dataset but failed on the DKASC-ASA-1B dataset, indicating architectural misalignment. Despite their good performance on the DKASC-ASA-2 dataset, Pyraformer and FEDformer had higher error rates than the PCA–Transformer model and Autoformer, indicating that the latter models are better at understanding and forecasting data dynamics. Autoformer was specifically designed for accurate predictions over longer periods using an auto-correlation mechanism. Meanwhile, Pyraformer is more suitable for handling long data sequences using its pyramidal attention module. FEDformer does not perform as well as the PCA–Transformer model but exhibits indications of error reduction because it uses frequency enhancement and transformer construction, enabling a comprehensive global time series forecasting. Table 3 presents a comparison of different Transformer-based models used for predicting wind and hydroelectric energy consumption. The models were evaluated using the MAE, RMSE, and R^2 metrics. With the lowest error rates (MAE and RMSE) and the most explanatory power(R^2), PCA-Transformer models in both 5-layer and 7-layer configurations often outperform other models, including classical ones like ARIMA and attentions one for both forms of energy. Significantly, when evaluating the models on datasets with varying domains and distributions, the PCA-Transformer consistently demonstrates exceptional performance. This highlights the strength and precision of energy forecasting, indicating that the combination of PCA and Transformers can significantly improve prediction abilities by handling disturbances and minimizing overfitting in intricate energy datasets. The models discussed herein employ distinct methodologies, including decomposition, efficient representation, and frequency analysis, to enhance the forecasting accuracy. When selecting a model, its quantitative performance measurements and qualitative design aspects of each approach must be considered in addition to the distinctive characteristics of the dataset and forecasting requirements. Figure. 4 shows the loss curve graph, wherein the PCA–Transformer model outperforms other transformer models during training epochs. The proposed model demonstrated efficient learning as the loss decreased rapidly, indicating optimal performance on DKASC-ASA-2 dataset.

Table 4 shows that the PCA–Transformer model outperforms state-of-the-art time series forecasting models in all aspects. The proposed model outperformed traditional ML algorithms such as SVM and extreme learning machine (ELM) and more recent neural network topologies such as DenseNet, RCC-LSTM, and ESNCNN on the DKASC dataset. The RCC-LSTM, which combines recurrent neural networks with LSTM, shows MAE and RMSE values of 0.587 and 0.94, respectively, indicating its low forecasting precision. DenseNet, characterized by its complex interconnections and depth, shows an MAE of 0.152; it does not provide RMSE, which is necessary for a comprehensive evaluation of its forecasting capabilities. SVM and ELM show MAEs of 0.2805 and 0.2367, respectively. ELM shows an RMSE of 0.2107. Although these models demonstrate high accuracy, they do not meet the standards established by more sophisticated models. The graph-based technique and ESNCNN show improvements; however, ESNCNN performs better with an MAE of 0.0971 and RMSE of 0.1731, exhibiting a nearly cutting-edge performance. The PCA–Transformer model outperforms these models by achieving notably lower MAE and RMSE values of 0.0074 and 0.11, respectively, thereby setting a new benchmark for forecasting accuracy. The outstanding performance of this model can be attributed to the combined use of PCA for extracting features and the ability of the transformer architecture to capture intricate temporal relationships. Thus, combining conventional dimensionality reduction methods with the sophisticated capabilities of contemporary neural

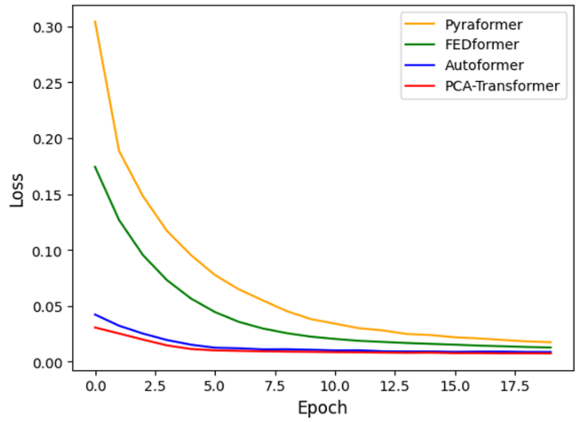


FIGURE 4. Loss curve graph exhibiting a distinct performance trend of the proposed model compared to various transformer models based on the DKASC-ASA-2 dataset.

TABLE 2. Comparative analysis of baseline transformer models with the PCA–transformer model for solar-power-generation prediction.

Method (baseline)	Dataset	MAE	RMSE	R^2
ARIMA	DKASC-ASA-2	1.06	1.94	0.90
Liu et al., 2022 (Pyraformer)	DKASC-ASA-2	0.0174	0.31	0.92
Zhou et al., 2022 (FEDformer)	DKASC-ASA-2	0.0126	0.29	0.94
Wu et al., 2022 (Autoformer)	DKASC-ASA-2	0.0087	0.33	0.91
PCA–Transformer (4 layer)	DKASC-ASA-2	0.0074	0.11	0.98
Liu et al., 2022 (Pyraformer)	DKASC-ASA-1B	0.43	0.77	0.89
Zhou et al., 2022 (FEDformer)	DKASC-ASA-1B	0.39	0.54	0.91
Wu et al., 2022 (Autoformer)	DKASC-ASA-1B	0.41	0.83	0.86
PCA–Transformer (7 layer)	DKASC-ASA-1B	0.12	0.2	0.95

TABLE 3. Comparison of the proposed PCA–transformer model with various transformer models on the EIA energy dataset.

Paper	Dataset	MAE	RMSE	R^2
ARIMA	EIA wind energy	3.7	7.62	0.96
Liu et al., 2022 (Pyraformer)	EIA wind energy	5.88	8.77	0.90
Zhou et al., 2022 (FEDformer)	EIA wind energy	5.21	8.11	0.92
Wu et al., 2022 (Autoformer)	EIA wind energy	6.12	9.02	0.89
PCA–Transformer (5 layer)	**EIA wind energy**	**4.23**	**7.29**	**0.97**
ARIMA	Hydroelectric Energy.	9.38	11.54	0.41
Liu et al., 2022 (Pyraformer)	Hydroelectric Energy.	7.92	8.77	0.69
Zhou et al., 2022 (FEDformer)	Hydroelectric Energy.	7.11	8.12	0.71
Wu et al., 2022 (Autoformer)	Hydroelectric Energy.	8.05	9.15	0.68
PCA–Transformer (7 layer)	**Hydroelectric Energy.**	**5.56**	**7.69**	**0.77**

TABLE 4. Comparison of the proposed PCA–transformer model with state-of-the-art methods [37] based on the DKASC datasets.

Paper	Method	Dataset	MAE	RMSE
Chen *et al.* (2020) [40]	RCC-LSTM	DKASC	0.587	0.94
Zang *et al.* (2020a) [41]	DenseNet	DKASC	0.152	-
Li *et al.* (2019) [42]	SVM	DKASC	0.2805	-
Zhou *et al.* (2020) [43]	ELM	DKASC	0.2367	0.2107
Cheng *et al.* (2021) [44]	Graph	DKASC	0.177	0.336
Khan *et al.* (2022) [37]	ESNCNN	DKASC	0.0971	0.1731
Ours	PCA–Transformer	DKASC	0.0074	0.11

networks is effective for substantially improving the time series prediction accuracy.

IV. CONCLUSION

The PCA-Transformer model exhibits outstanding accuracy in forecasting renewable energy production in diverse areas, such as solar, wind, and hydroelectric power. The model has demonstrated higher accuracy in both the DKASC datasets and regular EIA energy datasets by using PCA for accurate feature selection to build a transformer architecture to capture substantial temporal correlations. It demonstrates superior performance compared to well-established forecasting models such as Pyraformer, FEDformer, Autoformer, RCC-LSTM, and several machine learning methodologies including DenseNet, SVM, ELM, graph-based algorithms, and ESNCNN. The PCA-Transformer demonstrated superior performance in terms of accuracy (measured by MAE and RMSE) and explanatory power (measured by R^2) for solar, wind and hydroelectric energy. This confirms its resilience in handling diverse data distributions.

Moreover, the PCA-Transformer possesses the ability to adapt and modify its complexity in response to the underlying patterns in the data. This adaptability not only boosts the precision of forecasts but also improves the operational efficiency of the model, making it an attractive option for the many obstacles of renewable energy projection. The model showcased its wide range of uses by achieving outstanding results on the DKASC-ASA-2 dataset, with very low MAE and RMSE scores of 0.0074 and 0.11, respectively. Similarly, it performed well on the DKASC-ASA-1B dataset, with scores of 0.12 and 0.20 for MAE and RMSE, respectively.

The utilization of the PCA-Transformer has major economic ramifications as it improves the precision of renewable energy prediction, allowing for effective resource allocation and resulting in substantial cost reductions. These advantages enhance the stability of energy expenses and enhance the economic feasibility of renewable energy systems. The PCA-Transformer's ability to be replicated in various case studies has been highlighted, emphasizing its potential for wider influence and establishing it as a fundamental component for future developments in energy forecasting technology.Our future ambitions include improving the PCA-Transformer model and applying it to increasingly complicated time series prediction problems. This requires broadening real-time forecasting system applications and testing the model's interpretability and applicability outside of renewable energy. These projects aim to increase PCA-Transformer flexibility and predictive technology by improving prediction accuracy and addressing sustainability challenges. Clear explanations and future study goals are included in the revised language, underlining the model's potential impact.

AVAILABILITY OF DATA AND MATERIALS

Data is available at

1. https://dkasolarcentre.com.au/download?location = alice-springs
2. https://www.eia.gov/totalenergy/data/browser/?tbl = T10.01#/?f=M &start=198509&end=202403& charted=6-7-8-9-14

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] A. Q. Al-Shetwi, "Sustainable development of renewable energy integrated power sector: Trends, environmental impacts, and recent challenges," *Sci. Total Environ.*, vol. 822, May 2022, Art. no. 153645.
[2] A. Pani, S. S. Shirkole, and A. S. Mujumdar, "Importance of renewable energy in the fight against global climate change," *Drying Technol.*, vol. 40, no. 13, pp. 2581–2582, Oct. 2022.
[3] J. Zheng, J. Du, B. Wang, J. J. Klemeš, Q. Liao, and Y. Liang, "A hybrid framework for forecasting power generation of multiple renewable energy sources," *Renew. Sustain. Energy Rev.*, vol. 172, Feb. 2023, Art. no. 113046.
[4] Z. Li, J. Yu, G. Zhang, and L. Xu, "Dynamic spatio-temporal graph network with adaptive propagation mechanism for multivariate time series forecasting," *Exp. Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119374.
[5] C. Medina, C. R. M. Ana, and G. González, "Transmission grids to foster high penetration of large-scale variable renewable energy sources—A review of challenges, problems, and solutions," *Int. J. Renew. Energy Res. (IJRER)*, vol. 12, no. 1, pp. 146–169, 2022.
[6] V. Kumar, N. Kedam, K. V. Sharma, D. J. Mehta, and T. Caloiero, "Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models," *Water*, vol. 15, no. 14, p. 2572, Jul. 2023.
[7] O. Orang, P. C. L. E. Silva, and F. G. Guimarães, "Time series forecasting using fuzzy cognitive maps: A survey," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 7733–7794, Aug. 2023.
[8] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," *J. Econ. Surv.*, vol. 37, no. 1, pp. 76–111, 2023.
[9] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philos. Trans. Roy. Soc. A*, vol. 379, Apr. 2021, Art. no. 20200209.
[10] A. T. D. Perera, V. M. Nik, D. Chen, J.-L. Scartezzini, and T. Hong, "Quantifying the impacts of climate change and extreme climate events on energy systems," *Nature Energy*, vol. 5, no. 2, pp. 150–159, Feb. 2020.
[11] I. T. Jolliffe, *Principal Component Analysis for Special Types of Data*. Berlin, Germany: Springer, 2002.

[12] R. S. Tsay, *Analysis of Financial Time Series*. Hoboken, NJ, USA: Wiley, 2005.
[13] C. Alexander, "Market models: A guide to financial data analysis," Univ. Sussex, 2001. [Online]. Available: https://hdl.handle.net/10779/uos.23387939.v1
[14] J. Bai and S. Ng, "Large dimensional factor analysis," *Found. Trends Econ.*, vol. 3, no. 2, pp. 89–163, 2008.
[15] H. Lutkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer, 2005.
[16] R. Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography*. Amsterdam, The Netherlands: Elsevier, 1988, p. 425.
[17] J. Khan, E. Lee, A. S. Balobaid, and K. Kim, "A comprehensive review of conventional, machine leaning, and deep learning models for groundwater level (GWL) forecasting," *Appl. Sci.*, vol. 13, no. 4, p. 2743, Feb. 2023.
[18] S. Khan and H. Alghulaiakh, "ARIMA model for accurate time series stocks forecasting," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 524–528, 2020.
[19] A. K. Dubey, A. Kumar, V. García-Díaz, A. K. Sharma, and K. Kanhaiya, "Study and analysis of SARIMA and LSTM in forecasting time series data," *Sustain. Energy Technol. Assessments*, vol. 47, Oct. 2021, Art. no. 101474.
[20] D. M. Ahmed, M. M. Hassan, and R. J. Mstafa, "A review on deep sequential models for forecasting time series data," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, pp. 1–19, Jun. 2022.
[21] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," 2021, *arXiv:2104.05704*.
[22] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," 2022, *arXiv:2202.07125*.
[23] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," 2022, *arXiv:2211.14730*.
[24] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. X. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[25] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. NIPS*, vol. 34, Dec. 2021, pp. 22419–22430.
[26] J. Zhu, Y. Tan, R. Lin, J. Miao, X. Fan, Y. Zhu, P. Liang, J. Gong, and H. He, "Efficient self-attention mechanism and structural distilling model for Alzheimer's disease diagnosis," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105737.
[27] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 27268–27286.
[28] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–20.
[29] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artif. Intell. Med.*, vol. 104, Apr. 2020, Art. no. 101822.
[30] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106622.
[31] P. J. Pereira, C. Gonçalves, L. L. Nunes, P. Cortez, and A. Pilastri, "AI4CITY—An automated machine learning platform for smart cities," in *Proc. 38th ACM/SIGAPP Symp. Appl. Comput.*, Mar. 2023, pp. 886–889.
[32] A. Mosavi, M. Salimi, S. F. Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, p. 1301, Apr. 2019.
[33] R.-G. Cirstea, C. Guo, B. Yang, T. Kieu, X. Dong, and S. Pan, "Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting-full version," 2022, *arXiv:2204.13767*.
[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
[35] X. Suau, U. Zappella, and N. Apostoloff, "Filter distillation for network compression," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3129–3138.
[36] A. A. M. H. Asbahi, F. Z. Gang, W. Iqbal, Q. Abass, M. Mohsin, and R. Iram, "Novel approach of principal component analysis method to assess the national energy performance via energy trilemma index," *Energy Rep.*, vol. 5, pp. 704–713, Nov. 2019.
[37] Z. A. Khan, T. Hussain, I. U. Haq, F. U. M. Ullah, and S. W. Baik, "Towards efficient and effective renewable energy prediction via deep learning," *Energy Rep.*, vol. 8, pp. 10230–10243, Nov. 2022.
[38] EIA. (Jun. 2024). *USEIA EIA Total Energy*. [Online]. Available: https://www.eia.gov/totalenergy/data/browser/?tbl=T10.01#/?f=M&start=198509&end=202403&charted=6-7-8-9-14
[39] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2623–2631.
[40] B. Chen, P. Lin, Y. Lai, S. Cheng, Z. Chen, and L. Wu, "Very-short-term power prediction for PV power plants using a simple and effective RCC-LSTM model based on short term multivariate historical datasets," *Electronics*, vol. 9, no. 2, p. 289, Feb. 2020.
[41] H. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Wei, and G. Sun, "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning," *Int. J. Electr. Power Energy Syst.*, vol. 118, Jun. 2020, Art. no. 105790.
[42] L.-L. Li, S.-Y. Wen, M.-L. Tseng, and C.-S. Wang, "Renewable energy prediction: A novel short-term prediction model of photovoltaic output power," *J. Cleaner Prod.*, vol. 228, pp. 359–375, Aug. 2019.
[43] Y. Zhou, N. Zhou, L. Gong, and M. Jiang, "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine," *Energy*, vol. 204, Aug. 2020, Art. no. 117894.
[44] L. Cheng, H. Zang, T. Ding, Z. Wei, and G. Sun, "Multi-meteorological-factor-based graph modeling for photovoltaic power forecasting," *IEEE Trans. Sustain. Energy*, vol. 12, no. 3, pp. 1593–1603, Jul. 2021.

FAHMAN SAEED received the Ph.D. degree in computer science, specializing in the development of automated machine learning systems. He is currently an Assistant Professor with the College of Computer and Information Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), where he is a highly esteemed expert in artificial intelligence (AI). His contributions to optimizing AI model construction have led to increased precision and reduced resource consumption. In addition, his application of AI in medical imaging has enhanced diagnostic accuracy and informed healthcare decision-making. His innovative forecasting algorithms and models have proven highly accurate in predicting industry trends. His commitment to AI research and its practical applications demonstrates his ability to address real-world challenges and make lasting contributions to society. His extensive research in renewable energy has significantly impacted environmental conservation and sustainability.

SULTAN ALDERA received the master's and Ph.D. degrees from The University of Manchester, Manchester, U.K., in 2015 and 2020, respectively. He is currently an Assistant Professor in computer science. In 2015, he was a Researcher at AstraZeneca, U.K., applying machine learning techniques to gain insight into cancer clinical trials leading to cost reduction in terms of time, pharmaceutical, laboratory, and human resources. In 2020, his Ph.D. study contributed to introduce the ARQ framework, an innovative system for evaluating the reproducibility of medical AI research. In 2022, he took the helm of the Computer Science Department, guiding 120 faculty members and mentoring 1500 students across four programs, such as bachelor's and master's degrees in computer science, the master's degree in artificial intelligence, and the master's degree in information security. His expertise lies in artificial intelligence, bioinformatics, and natural language processing, with a keen focus on transparency and reproducibility.