RESEARCH ARTICLE

Weather Phenomena Monitoring: Optimizing Solar Irradiance Forecasting With Temporal Fusion Transformer

XINYANG HU

College of Art and Science, New York University, New York, NY 10003, USA

e-mail: huxinyang0813@163.com

ABSTRACT Global climate change has spurred the search for renewable energy sources, with solar power being a cost-effective option for electricity generation. Accurate energy generation forecasting is crucial for efficient usage planning. While various techniques have been introduced, transformer-based models are effective for capturing long-range dependencies in data. This study proposes an hour-ahead solar irradiance (SI) forecasting framework based on variational mode decomposition (VMD) for handling the meteorological data and modified temporal fusion transformers (TFT) for forecasting solar irradiance. The proposed model decomposes the raw solar irradiance sequence into intrinsic mode functions using VMD and optimizes the TFT using the variable screening network and GRU-based encoder-decoder. The resulting deep learning model provides interpretable outputs such as importance ordering of solar irradiance sub-sequences, and multi-head attention analysis of different forecasting window sizes. Empirical study shows that transformers have high performance on the long-term dependencies as compared to other time series models such as ANN and LSTM. On the e National Solar Radiation USA SI dataset, the proposed TFT achieved an MAE of 19.29 and an R^2 of 0.992 for a one-hour ahead forecast, while on the Pakistan SI dataset, it achieved an 11.46% reduction in MAE and a 7.12% improvement in MSE compared to the original TFT.

INDEX TERMS Renewable energy, transformers, temporal fusion transformers (TFT), energy forecasting, solar irradiance forecasting.

I. INTRODUCTION

Energy forecasting refers to predicting the proportion of energy produced either from renewable energy (RES; hydro, wind, solar) or non-renewable energy (natural gas, oil, coal) called fossil fuel sources, load prediction, etc. Forecasting energy-related tasks are of vast importance as secure and cost-effective transmission is a major need today. Forecasting has played an important role in the power and energy industry as well as it significantly in business decision-making [1]. The world is facing significant challenges associated with their forecasting, i.e., the intermittent and turbulent nature of these sources [2]. Solar energy, which is inherently variable, is affected by geographic and temporal factors, leading to

differing energy irradiance across locations—especially in regions with contrasting day lengths. This variability, coupled with the depletion of finite fossil fuel resources and the negative impacts of climate change, underscores the necessity for improved solar energy forecasting methods.

Recent studies have highlighted the efficacy of advanced machine learning techniques, particularly transformer-based models, for accurately predicting solar irradiance. The Temporal Fusion Transformer (TFT) has emerged as a prominent choice due to its ability to capture long-range dependencies and its interpretability, which is critical for understanding the underlying patterns in solar irradiance data. Notable works have demonstrated the effectiveness of TFT in various forecasting contexts, suggesting that its architecture is well-suited for the complexities of solar irradiance forecasting [3]. For instance, research has shown

---

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

that incorporating historical meteorological data enhances the accuracy of forecasts [4], a principle that we aim to leverage in our study.

Accurate forecasting can play a vital role in optimizing the utilization of renewable energy sources [5], as the integration of these sources with grid electricity has the potential to meet energy requirements more efficiently. Consequently, investments in renewable energy technologies have witnessed a significant surge, with many countries setting ambitious targets for deploying renewable energy to reduce greenhouse gas emissions and ensure energy security. To improve solar irradiance forecasting, this study proposes a TFT framework combined with Variational Mode Decomposition (VMD) for effectively processing historical solar irradiance data along with other meteorological variables. Previous research has identified the challenges of data noise and the benefits of decomposing solar irradiance time series into intrinsic mode functions for improved forecasting performance [6]. This paper builds upon these insights, aiming to enhance the accuracy and reliability of solar irradiance predictions in real-world applications.

The depletion of fossil fuels, the impact of climate change, and the need for energy independence have emerged as pressing global issues. To address these concerns, there has been a surge in demand for renewable energy sources such as solar and wind power [7]. As a major renewable energy resource, solar power is projected to become a primary power source in the future due to its abundance and minimal carbon emissions. In 2018 [8], there was an 8% increase in the overall capacity of renewable power sources such as Hydropower (increasing from 1112 GW to 1132 GW), Solar (increasing from 405 GW to 505 GW), and Geothermal Power (increasing from 12.8 GW to 13.3 GW). As of 2019 [9], the amount of worldwide solar energy is given in Figure 1.

Solar energy is an abundant source of energy that can be harnessed using solar PhotoVoltaic (PV) panels to generate electricity from sunlight [10]. Currently, households and industries utilize solar PV technology to produce electricity [11] as it offers the advantage of quick energy production during sunny weather or when there is sufficient sunlight. Additionally, installing solar PV systems is cost-effective [12] and convenient [13]. Moreover, solar energy is an unlimited source of energy. Nevertheless, the issue of producing electricity from solar photovoltaic (PV) panels persists, as various factors such as partial or complete shading from clouds resulting in reduced power generation [14], deterioration of capacitors or batteries [15], potential induced degradation [16], and uncontrollable environmental factors [17] may hinder the process. The amount of electricity produced from solar panels is directly proportional to global horizontal irradiance (GHI), which varies by time of day and location. The power output of solar panels is affected by meteorological conditions, resulting in intermittent production. Therefore, to integrate solar energy grids into existing infrastructures for grid stability and energy efficiency, it is crucial to have precise forecasts of GHI to account for

the intermittent nature of solar power. The short-term forecast is typically used to predict irradiation from one hour up to a week ahead, while the long-term forecast is utilized for predicting seasonal impacts on irradiation. Long-term forecasting is critical for financial planning and revenue generation, while short-term forecasting is crucial for managing utilities [18].

Much research has been dedicated to developing advanced solar PV panels that can boost the production of electrical energy and enhance flexibility [19]. In addition to fixed models, portable solar PV panels have gained prominence and are being utilized in a variety of devices, including embedded Internet of Things (IoT) devices [20], electrical appliances, and monitoring sensors. The integration of computer programming technology has emerged as a critical aspect of solar PV operations, enabling efficient management and control of solar PV systems for optimal performance [21]. Forecasting, a subset of computer programming technology, involves predicting significant events and analyzing future trends. For instance, forecasting can be utilized to predict the deterioration of solar PV panels or solar irradiance [22]. The ability to accurately forecast solar irradiance has a positive impact on the solar PV system, leading to economic benefits. The objective of this study is to enhance the accuracy and dependability of solar irradiance predictions even in instances of missing data. The occurrence of pyranometer malfunction resulting in missing solar irradiance readings is not uncommon and can be attributed to various factors such as turbulent weather, device failure, signal interference, and uncontrollable circumstances. In this work, we have presented an enhanced temporal fusion transformer framework (TFT) to improve the accuracy of mid-term hourly load time-series forecasting. The main contributions of our study are as follows:

- This research represents an innovative method of a temporal fusion transformer (TFT) model for effective and comprehensive solar irradiance forecasting, which is among the initial attempts in this field. Traditionally, deep learning techniques have struggled to deliver the interpretability of results, often at the expense of either reduced prediction accuracy or higher computational complexity. However, TFT overcomes these limitations by offering two key interpretive outcomes, enabling the identification of crucial variables for solar irradiance prediction and facilitating the analysis of persistent temporal patterns, while preserving forecast accuracy and avoiding excessive computational costs.

- This study also combines the Variational mode decomposition (VMD) with TFT algorithms to enhance the accuracy and stability of solar irradiance prediction. VMD facilitates the decomposition of raw data into multiple subsequences to streamline the task of forecasting solar irradiance. Empirical results show that VMD enhances TFT's performance compared to original-TFT, which establishes its reliability and stability.

Installed solar energy capacity, 2022

Cumulative installed solar capacity, measured in gigawatts (GW).

Our World in Data

0 GW   5 GW   25 GW   75 GW

No data   1 GW   10 GW   50 GW
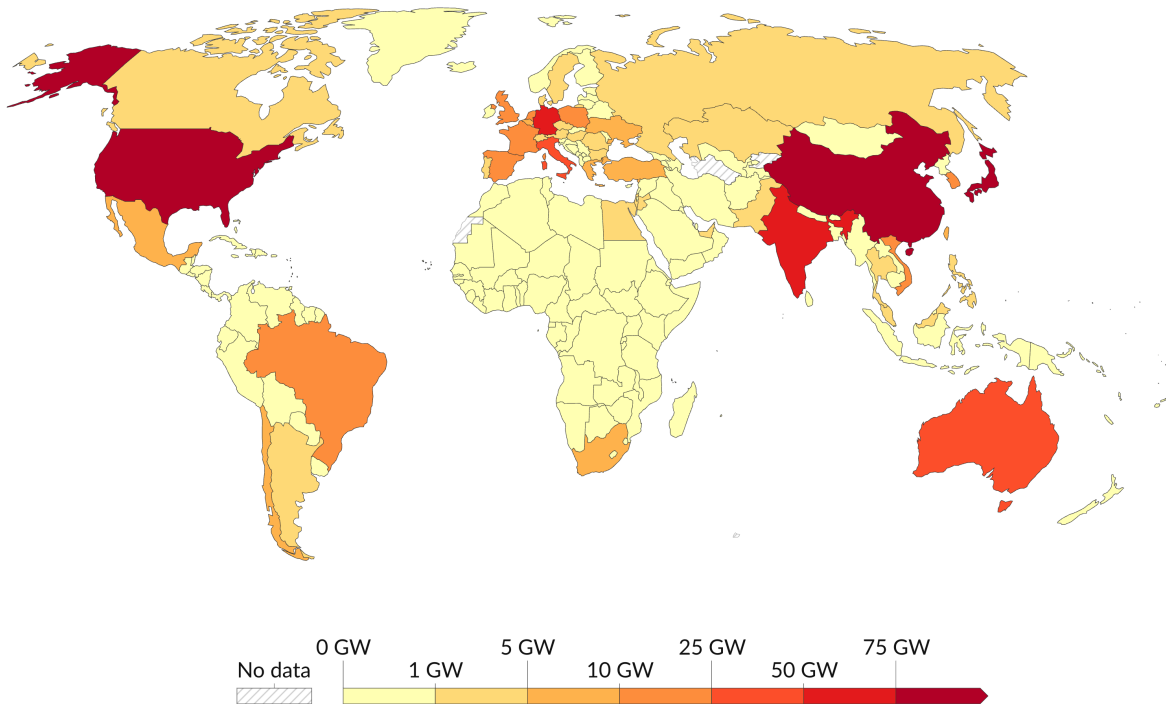
Data source: International Renewable Energy Agency (2023)   OurWorldInData.org/renewable-energy | CC BY

FIGURE 1. **Renewable energy generation by source, worldwide.**

- The Variable screen network and multi-horizons fusion interpretable multi-head attention module, remains a vital component of the TFT model. Its purpose is to provide a specific contribution of the individual input's impact on the output and to concentrate attention on the useful information dynamically. In contrast, the LSTM encoder-decoder is substituted with the GRU, which enables more effective learning of long-term temporal dependencies.

The paper is structured as follows: Section II discusses the overview of the existing research which contains a survey of prior work regarding solar irradiance, section III describes the problem formulation and methodology adopted to forecast the hour-ahead solar irradiance. Section IV describes the experiment setting and discusses the proposed model performance and then compares it with other start-of-the-art models trained on the same solar Irradiance dataset. In the end, Section V concludes the research with future direction.

II. LITERATURE REVIEW

In different areas, energy forecasting is a requirement as it is a key step for energy management. Numerous statistical and hybrid models have been conducted in recent years for energy generation and consumption forecasting using various state-of-the-art architectures. Energy forecasting is required in renewable energy plants and setups to fulfill the power requirements of systems where they are applied. Extensive research has been conducted over the last ten years to predict renewable energy production and electricity demand [24]. These predictions, with varying forecasting time frames such as short-term, medium-term, and long-term, serve various energy management purposes. Reference [25] utilized short-term forecasting for managing the energy flow among power sources, loads, and storage devices. On the other hand, medium-term and long-term forecasting primarily play a role in determining price settlements, load dispatch, and maintenance schedules. Reference [26] proposed a renewable power generation and electric load prediction approach using a stacked GRU-RNN based on AdaGrad and adjustable momentum. Firstly, a correlation coefficient was applied to input data for multiple sensitive monitoring parameters selection. Then selected parameters were inputted to stacked GRU-RNN for accurate RE generation or electricity load forecasting.

Several machine-learning models are used for residential and commercial energy forecasting [27]. Reference [28] proposed a technique for demand forecasting that involves

integrating the Bayesian probabilistic approach with deep learning methods. This approach produces prediction intervals (PIs) as a form of the forecast, indicating the range within which the forecast is likely to lie with a certain probability. In probabilistic deep learning, prediction intervals (PIs) are utilized to measure the level of uncertainty related to both the model parameters and energy consumption. This is in contrast to the conventional deep neural networks, which are deterministic and do not incorporate such uncertainties [29]. The [30] article discusses neural networks with a deterministic nature, as well as Bayesian neural networks (BNNs) which were suggested for solar irradiation forecasting. In [31] utilized Bayesian model averaging to predict the ensemble forecasting of solar power generation. Reference [32] explored BNNs for load forecasting and found that due to the probability distribution of model parameters, BNNs take longer to converge compared to other probabilistic deep learning methods, resulting in larger dimensions.

Deep learning methods, such as RNNs and LSTMs, have gained popularity for energy forecasting. Bidirectional-LSTM processes data bi-directionally and has been developed to capture patterns and improve accuracy in long-term energy forecasting, with potential benefits for energy management systems [33]. By uncovering inherent nonlinear features, deep learning has achieved exceptional performance in a range of applications such as energy production and load forecasting, as noted in reference [2]. Reference [34] presented a method for forecasting solar power output that relies on a deep feedforward neural network. This approach utilizes data from multiple sources, including solar power and weather forecast data. Reference [35] predicted solar radiation intensity based on several kinds of weather and solar radiation parameters.

An electricity forecasting model has been developed using a hybrid CNN-LSTM architecture [36]. The CNN component is responsible for feature extraction, while the LSTM layer handles the temporal characteristics of the time series data. Additionally, a predictive model for energy consumption has been proposed, which utilizes LSTM along with the sine cosine optimization algorithm [37]. Reference [38] suggests a hybrid approach that merges CNN and bi-directional Multilayer LSTM to create a sequential learning prediction model for energy usage. Additionally, another study introduces a similar hybrid model that combines CNN and Gated Recurrent Units (GRU) to improve the reliability of energy usage prediction. Both models utilize a coherent structure to achieve their respective goals. A similar model was presented by [36], in which a CNN-LSTM hybrid model was used for residential energy consumption forecasting. A hybrid architecture of CNN and ANN was proposed to get the advantage of both structures [39]. The proposed model was evaluated on RTE Power demand data and weather forecast dataset.

Numerous models have been extensively developed for time series forecasting owing to their great significance.

Classic tools such as [40] serve as the foundation for many time series forecasting techniques. ARIMA [41] addresses the forecasting challenge by converting the non-stationary process into a stationary one via differencing. Another approach for series forecasting is the filtering method [42]. In addition, RNN models have been employed to capture the temporal dependencies in time series [43]. DeepAR [44] is a forecasting technique that models the probability distribution of future series using a mix of autoregressive techniques and RNNs. To capture both short- and long-term temporal trends, LSTNet [45] uses CNNs with recurrent-skip connections. Temporal attention is used by attention-based RNNs [46] to analyze long-range dependencies and make predictions. Moreover, a number of research based on temporal convolution networks (TCN) [47] use causal convolution to model temporal causation. The primary emphasis of these advanced prediction models is on modeling temporal relationships using techniques such as recurrent connections, temporal attention, or causal convolution.

In recent times, there has been a surge of interest in Transformers [48] which leverage the self-attention mechanism to analyze sequential data, including natural language processing [49], audio processing [50], and computer vision [51]. Nonetheless, utilizing self-attention for forecasting long-term time series data poses a computational challenge due to the quadratic growth in both memory and time with respect to the sequence length L. The encoder-decoder framework, as implemented in the transformer model [52], employs an attention mechanism that allows the model to selectively attend to relevant information, addressing the shortcomings of the S2S model. This approach also enables parallel training, resulting in reduced training time. Reference [53] presents the TFT that employs both the transformer structure and quantile regression (QR) to learn temporal properties and generate probability predictions. For global solar radiation prediction, [54] presented a time series model. The author developed three models, one was used for daily forecasting and two for hourly forecasting (with and without night times). Predictions were performed using the nonlinear autoregressive models. Reference [55] presented a variant of the transformer-based model for electrical load forecasting by improving the NLP transformer model. The results demanded the likelihood of developing a pre-trained transformer model.

A machine learning-based new method with the stationary wavelet transform (SWT) and transformers was presented for household power consumption forecasting [56]. Experimental findings suggested that the proposed approach outperformed the existing approaches. One disadvantage was that the system could fail under unknown circumstances. A transformer-based energy forecasting model was developed using Pearson Correlation Coefficient analysis to boost prediction accuracy [57]. PV power output from North Taiwan, with one-hour intervals and a range of 2 years of data were used for the evaluation of the model. In [58],
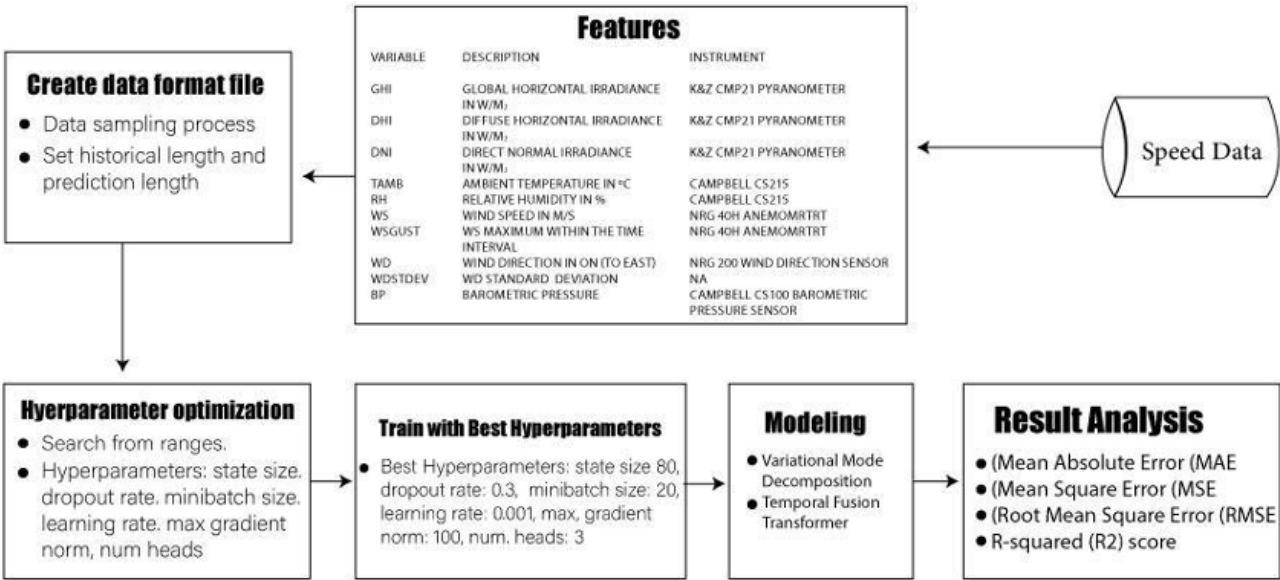
FIGURE 2. Workflow of Forecasting the solar Irradiance using temporal fusion transformer.

various techniques for forecasting wind speed have been proposed and utilized. By analyzing weather factors such as air pressure, humidity, and temperature, physical models are capable of forecasting energy [59], [60]. In the medium- and long-term, physical models are known to provide more accurate predictions. However, due to the absence of historical data trend characteristics, these models exhibit poor performance in the short term [61].

The initial step in predicting time-series data involves building a model based on Transformers [62]. This approach utilizes a self-attention mechanism to effectively capture intricate patterns and dynamics present in the time series dataset. Additionally, [63] introduces a Transformer network that delivers highly precise solar power generation forecasts, surpassing the performance of Linear Regression, CNN and LSTM methods. Reference [64] forecast short-term solar irradiance data using 13 datasets from various sources worldwide, with a maximum forecast horizon of 15. The Hidden Markov Model was extended to an infinite space dimension. Experimental results demonstrated that the proposed model produced more stable forecasting results for higher horizons compared to those obtained from the Markov-chain mixture distribution model. Reference [65] suggested utilizing the Particle Swarm Optimization Extreme Learning Machine (PSOELM) to forecast the short-term solar PV power in Beijing, China, and Qinghai. Their experiment yielded results indicating that PSOELM produced a significantly lower error compared to the traditional.

III. METHODOLOGY

This study uses the Variational Mode Decomposition in combination with the temporal fusion transformer to forecast

hour-ahead Solar Irradiance. TFT [62] was proposed for time series forecasting, and this work modifies some of its original components to enhance the robustness of forecasts on longer time horizons. The forecasting workflow is illustrated in Fig. 2. Workflow starts with raw data and extracts historical sequences and related metaphorical features. Then, creates the horizon windows on which the model takes the past sequence and forecasts the future. Hyperparameters are selected randomly and the dataset is split into an 80% train set and a 20% validation set. The modeling part consists of statistical component variational mode decomposition and modified temporal fusion transformer. Performance is analyzed using different performance metrics.

A. VARIATIONAL MODE DECOMPOSITION

VMD is an advanced signal decomposition technique that is adaptive and non-recursive. It can efficiently decompose non-stationary signals into a series of sub-patterns with robust decomposition performance, making it ideal for dealing with data noise. One of its major advantages is its fast and straightforward optimization process. The VMD method is utilized to extract multiple band-limited intrinsic modes concurrently from a complex signal. These intrinsic modes, known as intrinsic mode functions (IMFs), are AM-FM signals that represent the band-limited components of the original time series signal f(t).

The analytic signal of each mode $u_k(t)$ is computed by VMD using the Hilbert transform, which is then utilized to produce a one-sided frequency spectrum. The frequency spectrum of $u_k(t)$ is then mixed with an exponent corresponding to the center frequency, shifting the spectrum to the baseband. The bandwidth of the demodulated signal is
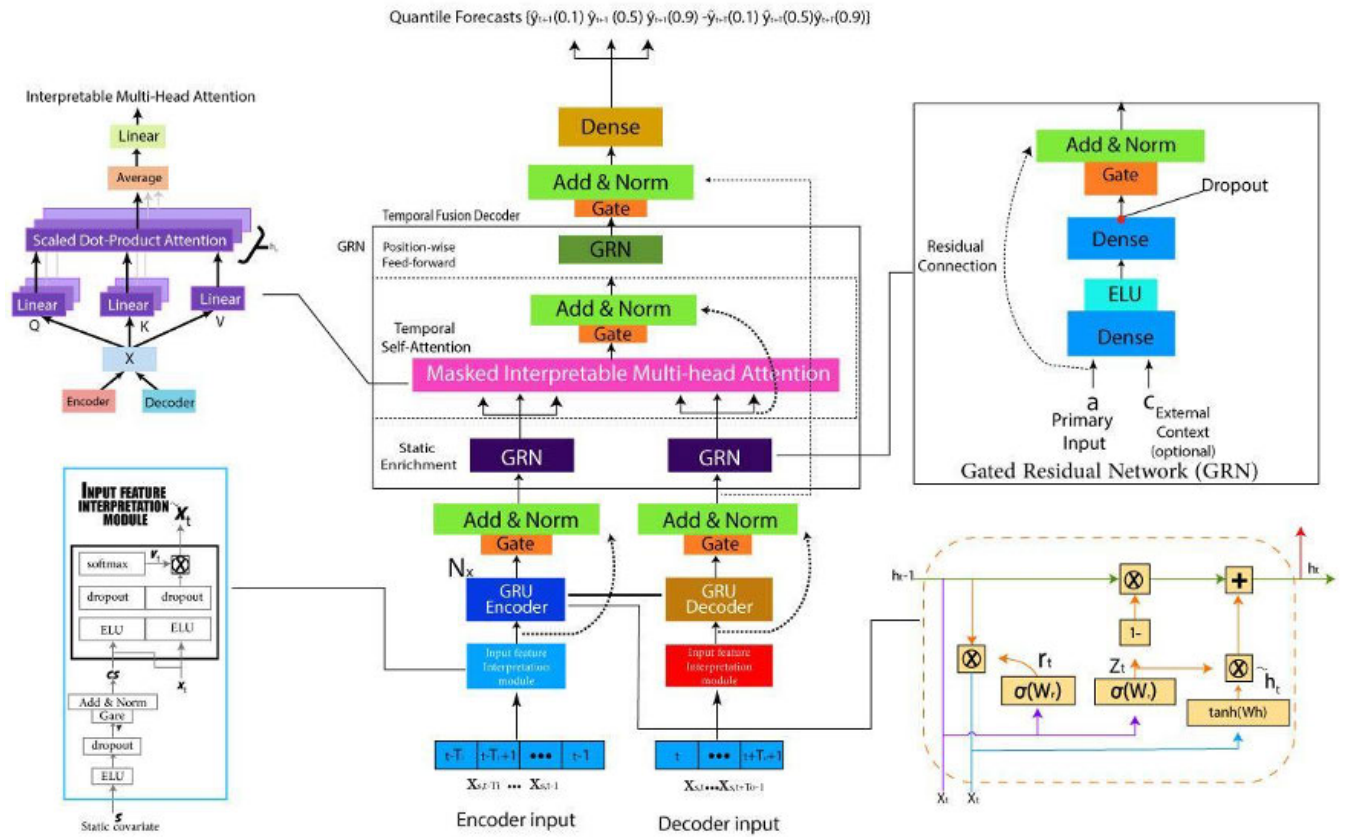
FIGURE 3. Framework of proposed temporal fusion transformer model.

determined using the H1 Gaussian smoothness, which is the squared $L2$-norm of the gradient. The primary objective of VMD is to decompose the raw signal $f(t)$ into multiple sub-modes ($u_k, k = 1, 2, \ldots, K$) that each has a center frequency ω_k. The decomposition is subject to a constraint condition that the sum of all the modes must be equal to the raw signal. VMD's objective function is to minimize the sum of each sub-mode frequency bandwidth that utilizes a constrained variational problem expressed as

$$VMD_{obj} = \text{Min}_{\{u_k, \omega_k\}} = \sum_{k=1}^{K} \left| \partial_t \left[\delta(t) + \frac{j}{\pi t} \right] * u_k(t) \right| e^{-j\omega_k t^2},$$

(1)

where u_k and ω_k are the decomposed modes and the corresponding center frequencies, respectively, where k ranges from 1 to K. Symbol $*$ denotes the convolutional operator, t represents time, $= j/2$, and δ represents the Dirac distribution. The decomposition is subject to the following constraint:

$$\sum_{k=1}^{K} u_k(t) = f(t)$$

(2)

The confined problem is equivalent to the unconstrained problem thanks to the use of α. On the other hand, alpha

ensures that the sub-model can be reconstructed accurately even in the presence of Gaussian noise. The unconstrained problem can be expressed as follows:

$$L(\{u_k, \omega_k\}, \lambda) = \alpha VMD_{obj} - \sum_{k=1}^{K} \|u_k(t)\|_2^2$$
$$+ \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t)$$

(3)

The alternate direction method of multipliers is used to solve the u_k and ω_k through Eqs (4) and (5).

$$u_k^{n+1}(\omega) = f(\omega) - \sum_{i \neq k} u_i(\omega) + \lambda(\omega)/2/1 + 2\alpha(\omega - \omega_k)^2$$

(4)

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k(\omega)|^2 d\omega}{\int_0^\infty |u_k(\omega)|^2 d\omega}$$

(5)

Eq (5) is used to update the center frequencies, ω_k, while equation (6) is used to update λ simultaneously. By utilizing the alternate direction method of multipliers as a foundation, the iteration stops once the termination condition, as given by

Eq. (7) and represented by ϵ, is achieved.

$$\lambda^{n+1}(\omega) = \lambda^{n+1}(\omega) + r(f(\omega) - \sum_{k=1}^{K} u_k^{n+1}(\omega)) \qquad (6)$$

$$\sum_{k=1}^{K} u_k^{n+1} - u_k^2/u_k^2 < \epsilon \qquad (7)$$

The real part of u_k^{n+1}, obtained through Fourier transform, is used to convert the final output of VMD, which is represented by u_k^{n+1}. The definition of r_{res} can be expressed as follows:

$$r_{\text{res}} = \frac{1}{N} \sum_{s=1}^{N} \left| \sum_{t=1}^{K} f(t) - \sum_{k=1}^{K} \frac{u_k(t)}{f(t)} \right| \qquad (8)$$

The number of IMFs K is considered optimal when the residual r_{res} does not exhibit a noticeable decreasing trend. In other words, if increasing K continues to lead to a significant reduction in r_{res}, it may indicate that additional modes are still capturing relevant information from the signal. Conversely, if increasing K yields a minimal change in r_{res}, it suggests that the decomposition has sufficiently captured the essential features of the original signal.

The raw wind speed data is represented by $f(t)$, with N denoting the sample number and K representing the number of decomposed sub-modes. The decomposed modes are represented by $u_k(t)$. To determine the number of patterns, r_{res} should not exhibit a noticeable decreasing trend [66].

B. TEMPORAL FUSION TRANSFORMER

Transformer-based models with an encoder and decoder have grown in popularity in time series forecasting [2]. The time series' past data is fed into the encoder, and the decoder uses autoregressive forecasting to predict future values, leveraging an attention mechanism to focus on the most valuable historical information for prediction. Nevertheless, past approaches frequently ignore various input types or presuppose that all external inputs will be known in the future. TFT, on the other hand, is a naturally interpretable model created especially for multi-horizon time series forecasting [34]. TFT addresses these issues by matching structures to certain data properties, providing better interpretability than black-box models like neural networks or complicated ensembles that conceal the relative relevance of features and their interactions. Furthermore, TFT represents a novel approach that surpasses other -state-of-the-art forecasting models.

In order to assure good prediction performance across various forecasting issues, TFT's model architecture, which is seen in Figure 3, uses established components to build feature representations for each input type (i.e., static, known future, and observed inputs). Separate inputs of the time-related input feature $Xs, t - Ti : t - 1$ and $Xs, t : t + To - 1$ with no shared parameters were made. Long-term time series' temporal information was effectively
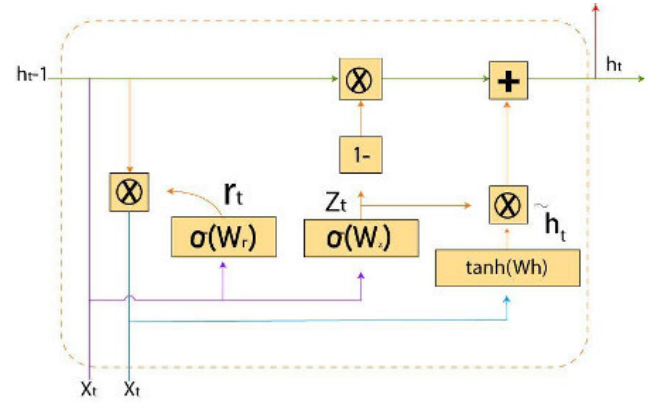
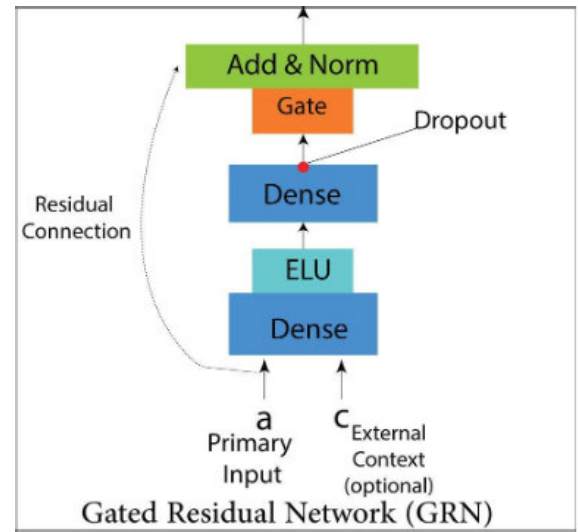

FIGURE 4. GRU structure.

FIGURE 5. Structure of GRN.

handled by multi-layer GRU encoders and decoders after the inputs, $\tilde{X}_{s,t-Ti:t-1}$ and $\tilde{X}_{s,t:t+To-1}$, had been transformed. The encoder and decoder outputs from the final layer were combined into a multi-timestep fusion module that distributed weights according to importance. By minimizing the revised quantile loss function, predicted quantile values were achieved. Note that modules with the same color in Figure 3 have the same parameters. TFT is made up of five main parts, which are broken down into gating mechanisms, variable screening networks, GRU encoder-decoder, interpretable multi-horizon fusion, and prediction intervals.

1) GATING MECHANISM

The gradient disappearance problem in RNN is resolved by GRU's use of temporal information processing, which keeps pertinent information and discards irrelevant ones. A reset gate and an update gate make up this system. Although the update gate stores memory from the previous time step, the reset gate merges fresh input data with previous memory.

Fig. 4 illustrates the GRU structure describes as follows:

$$GRN_\omega(a, c) = LN(a + GLU_\omega(\eta_1)) \qquad (9)$$

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega} \qquad (10)$$

$$\eta_2 = ELU(W_{2,\omega}p + W_{3,\omega}c + b_{2,\omega}) \qquad (11)$$

GRN is integrated as a block of TFT to provide the appropriate non-linear structure to the model. The input to GRN consists of two vectors: the primary information vector (p) and the context information vector (c). The computation of GRN is defined in Eq. (9), and the ELU activation function is defined in Eq. (11), where η_1 and η_2 are the outputs of two intermediate layers.

GLU plays a vital role in achieving gate control functionality. If the input is y, GLU's form is shown in Eq. (12), where $\sigma(.)$ is the sigmoid activation function, and $W(.)$ and $b(.)$ are the weights and biases. Illustrated in Fig. 5, GRN uses GLU to control the model structure and discard unnecessary layers. With GLU, GRN can regulate the model structure and avoid unnecessary layers.

$$GLU_\omega(y) = \sigma(W_{4,\omega}y + b_{4,\omega}) \odot (W_{5,\omega}y + b_{5,\omega}) \qquad (12)$$

2) VARIABLE SCREENING NETWORK

This work slightly modifies the original feature interpretation module [67] to identify the contribution of each input variable to the output for historic data. The module's structure is depicted in Fig. 6, which $X_t = \begin{bmatrix} x_{1,t} & x_{2,t} & \ldots & x_{k,t} \end{bmatrix}^T$ represents the input vector of k features at time t, and cs is the context vector obtained by the gated residual processing of the static covariate s.

$$\lambda = dropout(W_2(ELU(W_1s + b_1)) + b_2) \qquad (13)$$

$$O = LN(s + \sigma(W_3\lambda + b_3)) \odot (W_4\lambda + b_4) \qquad (14)$$

$$h_t = softmax(W_6(ELU(W_5(X_t) + b_5)) + b_6) \qquad (15)$$

$$\lambda = h_t \odot (dropout(W_8(ELU(W_1X_t + b_7)) + b_8)) \qquad (16)$$

The input feature interpretation module in Fig. 6 weighs input variables with weight vector h_t. The importance of each input feature is represented by h_t through the sigmoid activation function $\sigma(.)$, and the intermediate variable λ. The ELU activation function is used instead of RELU to alleviate gradient disappearance. The module includes dropout, layer normalization, and softmax layers. The larger the weight h_t, the more important the corresponding input variable is to the output.

3) INTERPRETABLE MULTI-TIMESTEP FUSION

A multi-timestep fusion module is designed to capture long-term linkages in hourly load forecasting. It produces a vector with the expected size by allocating weights to all of the encoders' and decoders' outputs from prior time steps. This enables the model to concentrate on more crucial data. Figure 7 depicts the multi-head fusion module's structure.

To improve interpretability and record long-term correlations between time steps, TFT makes use of self-attention. Values V are scaled by the modified transformer-based
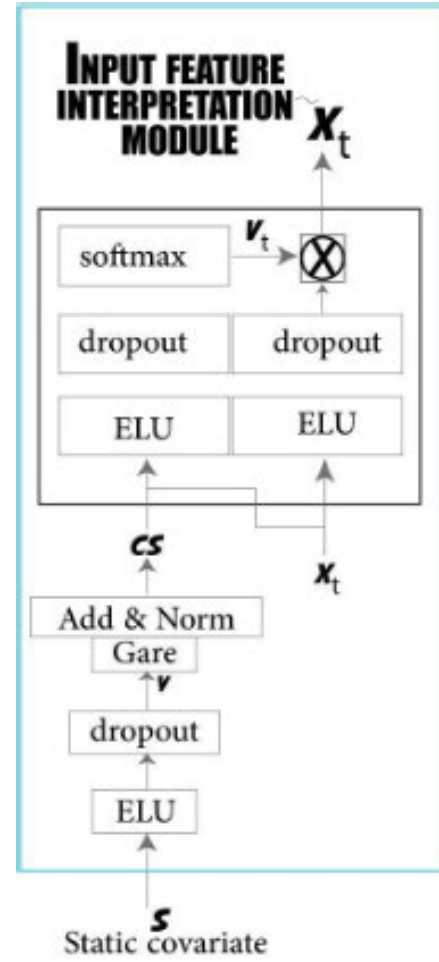

FIGURE 6. Structure of Variable Screening network.

multi-head attention mechanism according to the connections between questions Q and keys K. A multi-head attention structure is used by the TFT to learn long-term time series correlations. Eq. (17) is a definition of the self-attention mechanism, where Q, K, and V stand for query, key, and value, respectively. With the relationships between queries $Q \in \mathbb{R}^{N \times d_{att}}$ and keys $K \in \mathbb{R}^{N \times d_{att}}$ as a basis, attention mechanisms scale values as follows:

$$Attention(Q, K, V) = A(Q, K)V \qquad (17)$$

The linear transformation on the matrices Q, K, and V used in the self-attention mechanism is obtained by multiplying the input matrix X by the corresponding linear transformation matrices W_q, W_k, and W_v. The output matrix of the self-attention mechanism is then determined using these matrices. The typical multi-head attention technique, repeated self-attention, can be used to create several sets of Q, K, and V matrices. For attention values, scaled dot production is calculated as:

$$A(Q, K) = softmax(QK^T / \sqrt{d_{att}}) \qquad (18)$$
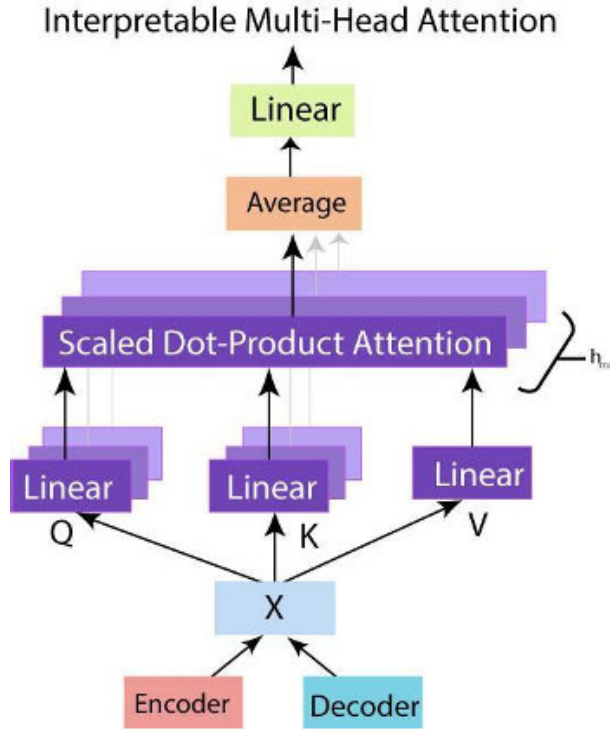
Interpretable Multi-Head Attention

FIGURE 7. Structure of multi-head feature fusion attention mechanism.

Proposed Multi-head distributes the matrix V among each head in order to address this. This makes it possible to assess the overall significance of a feature, and h_m multi-head attention is required, where h_m is the number of heads and W^h differs for each head. Multi-head attention uses several heads for various representation subspaces in order to increase the learning capacity of the attention mechanism.

$$InterpretableMultiHead(Q, K, V) = H_{1,H}W_H \quad (19)$$

$$H_h = Attention(QW_q^{(h)}q', KW_k^{(h)}K', VW_v^{(h)}V') \quad (20)$$

$[W_q^{(h)}q'W_k^{(h)}K']$ and $W_v^{(h)}V \in \mathbb{R}^{d_{model} \times d_{att}}$ refers to head-specific weights. As each head employs unique values, attention weights alone cannot accurately signify the significance of a specific feature.

$$H = Attention(Q, K)VW_v^{(h)}V' \quad (21)$$

$$= \{1/h_m \sum_{h_m=1}^{h_m} A(QW_q^{(h)}q'KW_k^{(h)}K')VW_v^{(h)}V'\} \quad (22)$$

$$= 1/h_m Attention(QW_q^{(h)}q'KW_k^{(h)}K')VW_v^{(h)}V'$$
$$h = 1, \ldots, h_m \quad (23)$$

The equation above depicts the multi-head attention mechanism, where $A(\cdot)$ represents a normalized function and n is the dimension of the K vector. The weight matrices of Q and K for the hth head are denoted by W_Q^h and KW_K^h, respectively, while W_V represents the weight matrix of V shared across all heads to enforce time-series causality.

C. LOSS FUNCTION
TFT predicts point predictions and quantile intervals simultaneously, where three percentiles 10th, 50th, and 90th are predicted at each step using the temporal fusion decoder's linear transformation. Quantile loss is jointly minimized during training. The outputs of all quantiles are subsequently summed up using the formula:

$$L(X, W) = \sum_{y \in X} \sum_{q \in \mathbb{Q}} \sum_{T=1}^{t_{max}} (QL(y_T, \hat{y}(q, t - T, T)q)/MT_{max})$$
$$(24)$$

$$QL(y, \hat{y}, q) = q(y - \hat{y})_+ + (1 - q)(\hat{y} - y)_+ \quad (25)$$

The weight parameters of TFT W are optimized by jointly minimizing the quantile loss over the set of output quantiles Q and the training data domain U, which consists of M samples. The function $(\cdot)_+$ denotes $\max(0, \cdot)$.

IV. EXPERIMENT AND RESULTS
In this section, we will explain the dataset and the process through which we have conducted our experimentation.

A. DATASETS
1) USA SI DATASET
The historical weather data utilized in this study was sourced from the National Solar Radiation Database (NSRDB), maintained by the National Renewable Energy Laboratory (NREL) under the United States Department of Energy [75]. The NSRDB provides a comprehensive collection of meteorological data, encompassing global locations and offering various temporal resolutions, including 15-minute, 30-minute, and 60-minute intervals. Users can select specific meteorological features based on their research needs. Data accessibility is facilitated through multiple channels, including the NSRDB Viewer, an application programming interface (API), and Amazon Web Services (AWS) cloud platforms. For this study, USA irradiance data recorded at one-hour intervals from 2018 to 2022 is used. These datasets include hourly data from 24 sites across three states in the United States, specifically Oklahoma, Texas, and Oregon. Data from 2018 to 2021 is used for model training, while the remaining data from January 1, 2021, to December 31, 2022, is utilized for model evaluation. This dataset includes not only irradiance values but also detailed date and time information, temperature readings, and other relevant weather parameters.

2) PAKISTAN SI DATASET
In this study, we have used a solar energy irradiance dataset from Islamabad [41], Pakistan, which contains various parameters related to PV generation, such as Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (DHI), and Direct Normal Irradiance (DNI), as well as wind speed and other data. This dataset was collected over five years, from 2016 to 2020, at an EMAP Tier 1 Meteorological Station at 33.64o North and 72.98o East, 500 meters above

sea level. The dataset consists of hourly records, and around 41 thousand records are available for analysis. By examining the data and analyzing the various spatial and temporal factors that impact PV efficiency, the researchers sought to better understand how to optimize solar energy generation in this region.

B. DATA SPLIT AND PRE-PROCESSING

Before feeding the data into each of the test models, we performed a data check to identify any missing values and the quantity of missing data. Additionally, during exploratory data analysis, a significant number of Global Horizontal Irradiance (GHI) values are equal to zero. These zero values corresponded to night-time solar GHI values and were expected to be zero. However, this large number of zero values impacted the accuracy of the models as there are 19403 GHI values, while there are around 41 thousand total records. To address this issue, all values between 7 PM and 5 AM, the typical nighttime hours, are eliminated from the dataset, reducing the number of zero values to 4236. After selecting data for a specific time interval, the total number of records available for analysis was 25785. This process has ensured transparency in the dataset being used for the modeling, thus improving the reliability of the results obtained from the models. The dataset was split into two subsets: a training set consisting of 70% of the data, and a validation set consisting of 30% of the data.

C. TRAINING AND VALIDATION STEP

For time series forecasting, a window is defined which consists of two values, the first one is the number of time steps taken to forecast the future value and the second one is the future values that are to be forecasted. Our experiments involved the use of four forecasting windows, each with a specific size. These windows were (7, 1), (14,3), (28, 7), and (56, 14), where the first number represents the number of past sequences and the second number represents the number of future sequences that the model needs to predict. Since the data sequences are hourly based, the model takes a historical data sequence of a specific number of hours and also forecasts future values for certain hours. For instance, in the first window size tuple (7, 1), the model takes a past sequence of 7 hours to predict the next hour's sequence. Similarly, the model takes past sequences of 14, 28, and 56 hours, and predicts 3, 7, and 14 hours into the future, respectively. We excluded zero-value transactions from the data between 7 pm and 5 am, leaving us with 14 hours of data sequences. Therefore, a day consists of 14 hours. Hence, forecasting using window sizes (7, 1) and (7, 3) means predicting solar irradiance based on half-day sequence, while forecasting using window sizes (28, 7) and (56, 14) involves taking 2 and 4-day past sequences to forecast next half and full-day solar irradiance, respectively. Our best results are recorded based on the second-mentioned window. Different experiments were conducted with different hyper-parameters and the best model was used for testing

purposes. We experimented with different values, batch sizes i.e., 16, 24, and 32, and learning rates i.e., 0.0001-0.1. The best results were obtained at a learning rate of 0.1 and batch size of 24. Quantile loss with seven quantiles was used as a loss function. Models are implemented on the Ubuntu system using Python 3.8 with TensorFlow, PyTorch-forecasting, PyTorch-lightning, and Torch. To split the data, the ''TimeSeriesDataSet'' Python library is utilized. The TFT model uses an ADAM optimizer and early stopping is employed to prevent overfitting. Training for the TFT model is performed on the GPU 1080 Ti having (R) Core (TM) i7 and 128 GB RAM.

D. EVALUATION METRICS

The following four evaluation metrics are used to evaluate our results for forecasting:

1) MEAN ABSOLUTE ERROR (MAE)

MAE is a commonly used metric for evaluating forecasting models. It is a measure of the difference between predicted values and actual values. Its mathematical form is given as follows:

$$MAE = \frac{1}{k} \sum_{t=0}^{k} |y_t - \hat{y}_t| \tag{26}$$

where y_t and $\hat{y}_t$ represent observed and predicted values, respectively.

2) MEAN SQUARE ERROR (MSE)

MSE is a measure of the difference between two continuous variables. It is a commonly used loss function for forecasting problems and is the average of the squares of the differences between the predicted and actual values. MSE is a way to measure the average magnitude of the error. It punishes larger errors more than smaller ones, which can be useful in certain types of problems. It is defined as:

$$MSE = \frac{1}{k} \sum_{t=0}^{k} (y_t - \hat{y}_t)^2 \tag{27}$$

3) ROOT MEAN SQUARE ERROR (RMSE)

RMSE is calculated by taking the square root of the MSE error value. It is also used as a forecasting metric and defined as below:

$$RMSE = \sqrt{\frac{1}{k} \sum_{t=0}^{k} (y_t - \hat{y}_t)^2} \tag{28}$$

4) COEFFICIENT OF DETERMINATION (R2) SCORE

R2 score measures how much of the dependent variable's variance is predictable by the independent variable(s) with a range of 0-1, where higher values indicate a better fit. The equation for R2 is:

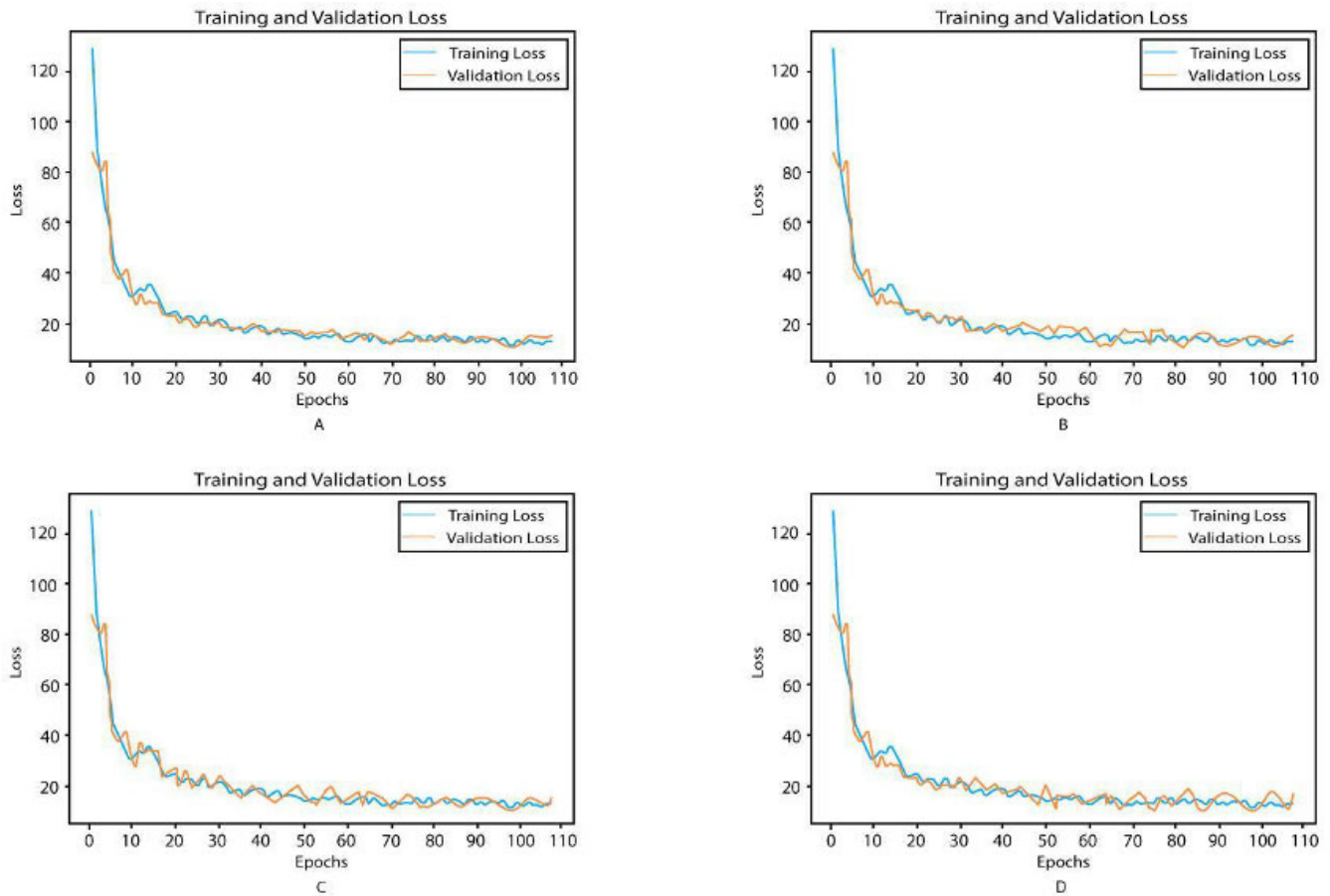$$R2 = 1 - \frac{SSE}{SST} \tag{29}$$

FIGURE 8. Loss graph of our modified TFT against horizon forecasting window size (7, 1), (7, 3), (28, 7), and (56, 14).

where the Sum of Squared Errors (SSE) is the sum of the squared deviation between the predicted and the actual outcomes and the Total Sum of Squares (SST) is the sum of the squared deviation between the actual values and the mean of the actual outcomes.

E. COMPARATIVE MODELS
1) ARTIFICIAL NEURAL NETWORKS (ANN)
ANN [76] is a machine learning model miming human information processing. ANNs consist of input, hidden, and output layers, with the hidden layer learning about data patterns through neurons. The hidden layer then forwards the result to the output layer. ANNs have self-learning capabilities and rely on data experiences to learn and predict results. This experiment uses the 2 hidden layers ANN with 64 and 32 hidden units, and one dense layer for output having 1, 3, 7 and 14 neurons for window size (7, 1), (14,3), (28, 7) and (56, 14), respectively.

2) LONG SHORT-TERM MEMORY (LSTM)
LSTM, an improved Recurrent Neural Network (RNN), is widely used for time series forecasting, and also used for solar Irradiance [41]. This study uses the 3 layers of LSTM with 64, 32, and 16 hidden units, and one dense layer

having 1, 3, 7, and 14 neurons for windows size (7, 1), (14,3), (28, 7) and (56, 14), respectively.

3) ORIGINAL TFT
Transformer-based models, comprising encoder and decoder parts connected via an attention mechanism, are famous for time series forecasting. The encoder takes historical data as input, while the decoder predicts future values by focusing on valuable historical information. The decoder uses masked self-attention to prevent future value acquisition during training. However, previous methods don't consider different input types or assume all exogenous inputs are known in the future. Original-TFT [74] addresses these issues by aligning architectures with unique data characteristics via appropriate inductive biases. In this work, the original TFT is trained for all 4 sizes of windows.

F. RESULTS AND DISCUSSION
This section analyses the performance of proposal TFT, LSTM, ANN, and original-TFT models with varying forecasting window tuples: (7,1), (14,3), (28,7), and (56,14). The first value in the tuple represents the number of past time steps used for forecasting where each step is an hour, while
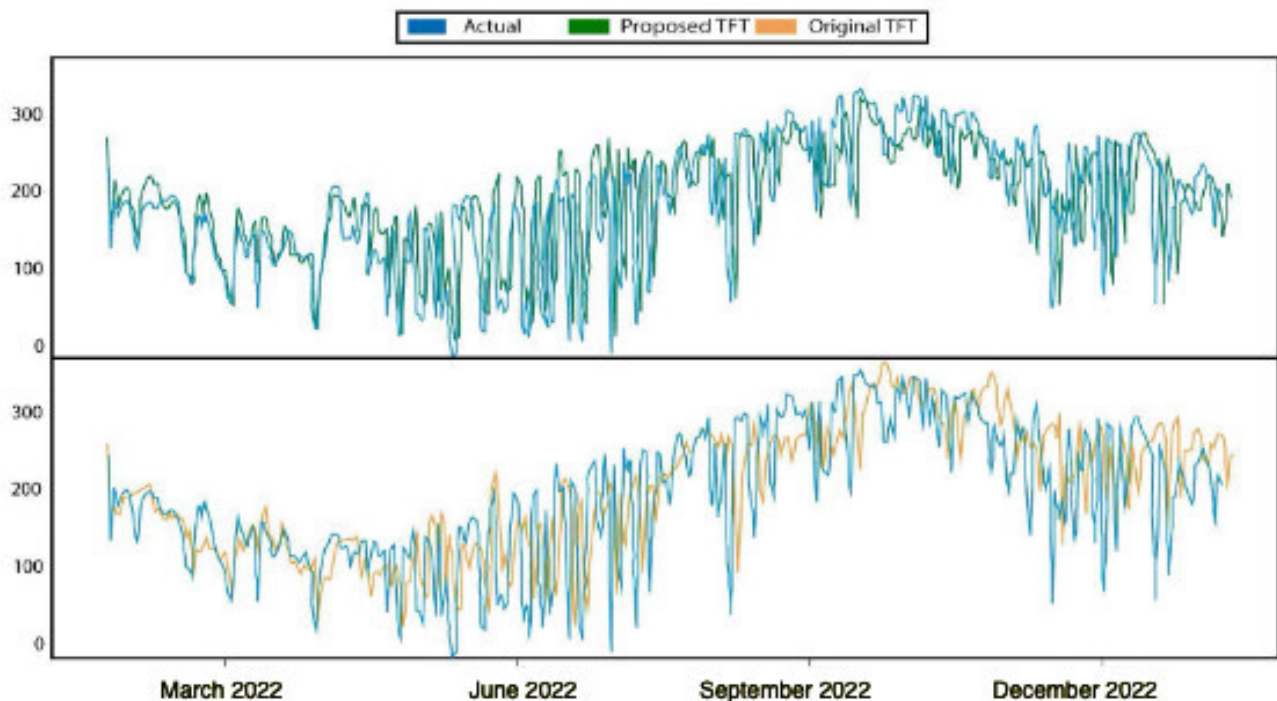
FIGURE 9. Forecast value against the actual value (above graph) mentioned the actual vs forecasted value against the proposed TFT, while the below graph mentioned the actual vs forecasted against the original TFT model.

the second value represents the number of time steps being forecasted which is also an hour long. A window size of (4,1) is a single-step forecast, while the others are multi-step forecasts ranging from 3 to 14 hours ahead. As the dataset has zero values of GHI from 7 PM and 5 AM, the solar day becomes 14 hours. Hence, window size (56, 14) means taking past 4 day data and forecasting next day data. The dataset used [41] in this study contains GHI data as well as weather data in which temperature, wind speed, etc. are variable quantities and metaphorical features.

1) TRAINING AND EVALUATION PERFORMANCE ON USA IS DATASET

This section presents the training and evaluation performance on the train and validation set, respectively. Training and validation loss plots of the proposed TF model are given in Figure 8 (A-B) against window size (7, 1) and (56, 14) on the USA SI dataset, respectively. Loss graphs show that performance starts decreasing on lengthening the forecasting window size. However, this deterioration in validation does not occur abruptly going from forecasting horizon (7, 1) to (14, 3), but sharp performance degradation can be seen from (28, 7) to (56, 14).

2) PERFORMANCE COMPARISON USA DATASET

This section presents a comparative performance analysis of the proposed Temporal Fusion Transformer (TFT), original TFT, Long Short-Term Memory (LSTM), and Artificial

Neural Network (ANN) models across four performance metrics: Mean Absolute Error (MAE), R^2, Mean Squared Error (MSE), and Root Mean Square Error (RMSE) for various forecasting window sizes: (7,1), (14,3), (28,7), and (56,14). For the one-hour forecast horizon (7,1), the proposed model significantly outperformed the others, achieving an MAE of 19.29 and an R^2 of 0.992. The ANN also surpassed the original TFT and LSTM, with MAE values of 21.53 and R^2 of 0.981, compared to 22.81 (MAE) and 0.971 (R^2) for original TFT, and 25.21 (MAE) and 0.926 (R^2) for LSTM. For three-hour forecasts (14,3), the proposed TFT and original TFT again performed better than LSTM and ANN, recording MAE values of 22.55 and 24.28, with R^2 scores of 0.978 and 0.960, respectively. As the forecasting horizon increases, the effectiveness of ANN and LSTM diminishes, as indicated by their R^2 scores, which decline from 0.882 and 0.831 (for (28,7)) to 0.835 and 0.818 (for (56,14)). The proposed TFT maintained R^2 scores of 0.940 and 0.921 for (28,7) and (56,14), respectively, while original TFT recorded 0.905 and 0.892.

The observed decline in accuracy with longer forecasting horizons can be attributed to increased complexity in prediction, the accumulation of errors, and greater data variability. Consequently, it is essential to consider these factors when evaluating the effectiveness of deep learning models for forecasting. The findings indicate that while these models are effective for short-term predictions, their reliability decreases over extended periods, necessitating

TABLE 1. A comparison of the proposed method with the state-of-the-art methods on USA SI dataset.

| Window	Model	MAE	MSE	RMSE	R2
(7,1)	LSTM	25.21	1619.70	40.24	0.92
	ANN	21.53	1251.51	35.37	0.98
	Original-TFT	22.81	1048.81	36.72	0.971
	Proposed TFT	19.29	957.10	30.93	0.99
(14,3)	LSTM	27.74	1883.52	43.39	0.90
	ANN	29.34	1937.99	44.02	0.895
	Original-TFT	24.28	1384.27	37.20	0.96
	Proposed TFT	22.55	1242.34	35.24	0.978
(28,7)	LSTM	79.20	5624.07	74.99	0.88
	ANN	95.84	7850.51	88.60	0.831
	Original-TFT	46.64	4918.68	70.10	0.905
	Proposed TFT	41.08	3742.34	61.17	0.94
(56,14)	LSTM	94.60	8606.80	92.77	0.835
	ANN	129.71	11029.39	105.02	0.818
	Original-TFT	77.91	6842.30	82.71	0.892
	Proposed TFT	63.51	5280.44	72.66	0.92
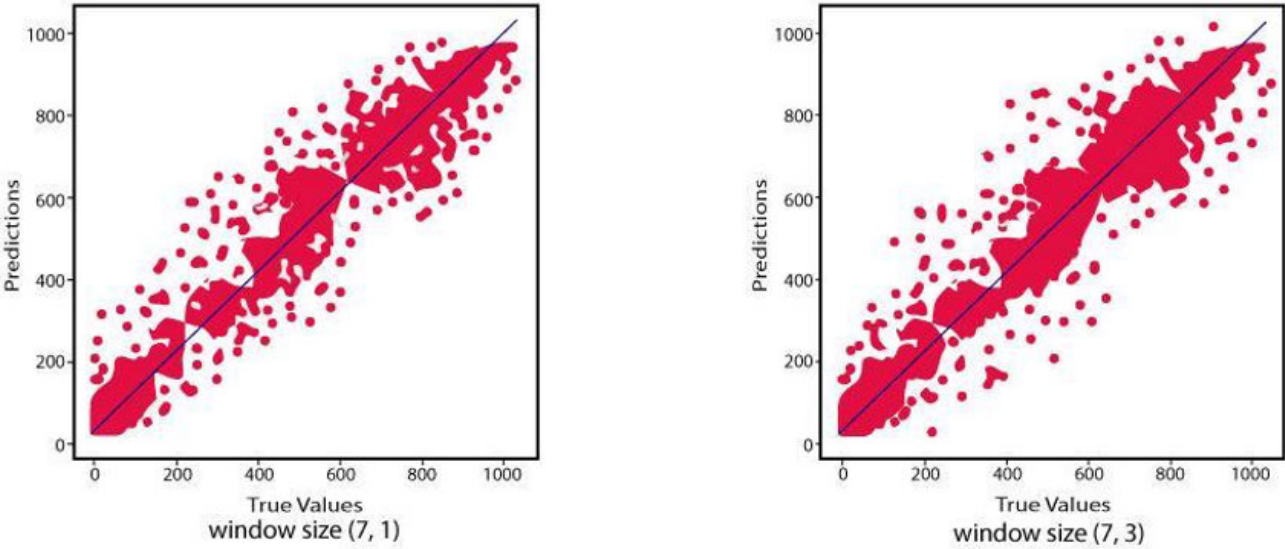
FIGURE 10. Scatter plots for proposed TFT on USA SI dataset.

careful selection of forecasting models based on the desired prediction horizon.

3) PERFORMANCE COMPARISON PAKISTAN DATASET
The proposed-TFT model demonstrates notable performance enhancements on the Pakistan SI dataset, achieving an 11.46% reduction in Mean Absolute Error (MAE) and a 7.12% improvement in Mean Squared Error (MSE) compared to the Original-TFT for a window size of (7,1). Additionally, it exhibits a 3.59% reduction in Root Mean Square Error (RMSE), confirming its efficacy in accurately predicting outcomes with shorter input sequences. The model's R^2 score of 0.99 indicates that it explains 99% of the variance in the data, which is critical for reliable forecasting.

As the window size increases to (14,3), the proposed-TFT maintains its superior performance, achieving an 8.01% reduction in MAE, an 8.29% reduction in MSE, and a 5.27% improvement in RMSE relative to the Original-TFT. The advantages become even more pronounced with larger window sizes. For (28,7), the proposed-TFT realizes a 5.46% reduction in MAE, a 22.70% reduction in MSE, and an 11.65% reduction in RMSE, demonstrating its robustness in managing increased complexity and extended prediction horizons. A high R^2 score of 0.98 further substantiates the model's reliability in complex forecasting tasks.

At the largest window size of (56,14), the proposed-TFT exhibits a 13.88% reduction in MAE, a 21.58% reduction in MSE, and a 9.36% reduction in RMSE compared to
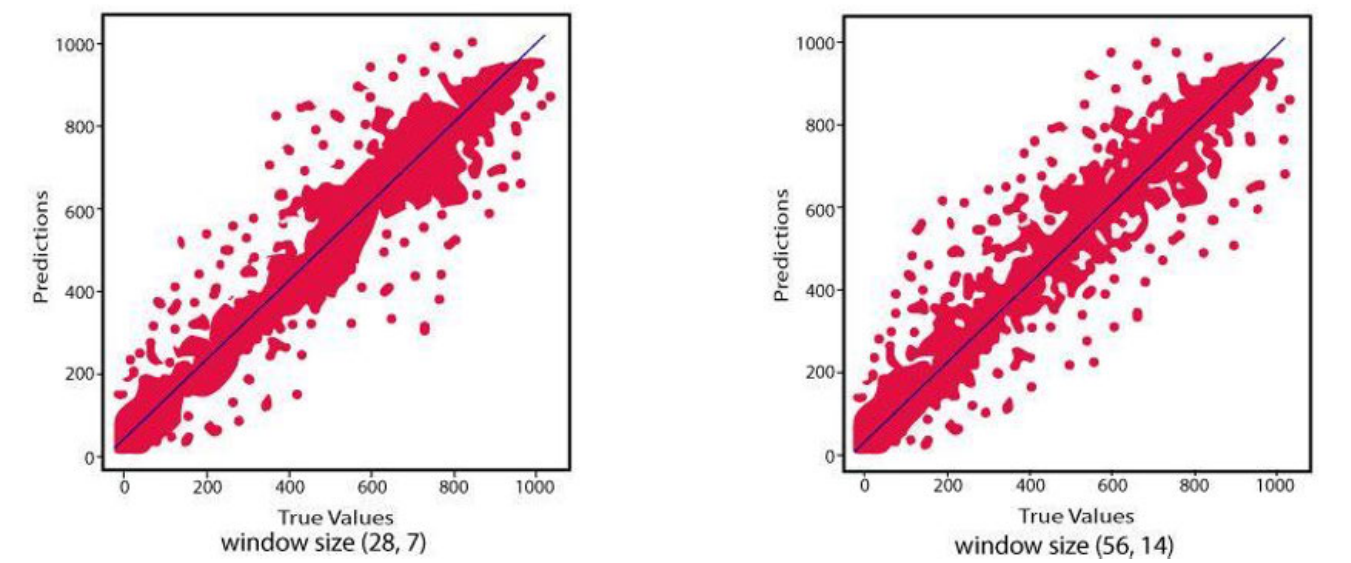
FIGURE 11. Scatter plots for proposed TFT on USA SI dataset.

TABLE 2. A comparison of the proposed method with the state-of-the-art method on Pakistan SI dataset.

Window	Model	MAE	MSE	RMSE	R2
(7,1)	LSTM	55.51	1939.65	44.04	0.93
	ANN	52.32	1591.37	39.89	0.96
	Original-TFT	53.37	1376.51	37.10	0.99
	Proposed TFT	49.14	1279.78	35.77	0.99
(14,3)	LSTM	58.88	2215.32	47.07	0.96
	ANN	60.10	2260.54	47.55	0.93
	Original-TFT	55.67	1718.53	41.46	0.96
	Proposed TFT	51.39	1566.40	39.58	0.98
(28,7)	LSTM	109.55	5955.44	77.17	0.97
	ANN	126.06	8190.14	90.50	0.96
	Original-TFT	75.86	5258.52	72.52	0.98
	Proposed TFT	71.71	4063.86	63.75	0.98
(56,14)	LSTM	123.40	8928.45	94.49	0.95
	ANN	159.06	11355.28	106.56	0.93
	Original-TFT	106.97	7162.11	84.63	0.95
	Proposed TFT	92.99	5619.03	74.96	0.97

the Original-TFT. Overall, the analysis across all window sizes consistently underscores the significant improvements provided by the proposed-TFT, achieving lower error rates and higher R^2 scores. These enhancements, ranging from 5.46% to 22.70%, affirm the proposed-TFT's capability to elevate forecasting performance, establishing it as an effective tool for accurate and consistent predictions across various temporal scales.

4) USA SI VS PAKISTAN SI DATASET COMPARSION

When window size increases from (7,1) to (56,14), the error metrics (MAE, MSE, RMSE) generally increase across all models, indicating a decline in performance. This degradation

is more pronounced in the Pakistan SI dataset, which suggests that larger window sizes introduce greater complexity and pose challenges in maintaining high accuracy. The increase in prediction horizon associated with larger window sizes likely contributes to this trend, as longer horizons are inherently more difficult to forecast accurately.

When comparing the results across the two datasets, it is clear that the models generally perform better on the USA SI dataset than on the Pakistan SI dataset. For instance, the MAE, MSE, and RMSE values are consistently lower on the USA dataset, and the R2 scores are higher, particularly for smaller window sizes like (7,1) and (14,3). This disparity could be attributed to several factors, including the quality and variability of the data in each dataset. The USA SI dataset

may have less variability or noise compared to the Pakistan SI dataset, making it easier for the models to learn and predict accurately.

The proposed-TFT model consistently shows significant improvements over the original-TFT model across both datasets and all window sizes. The percentage improvement in MAE, MSE, and RMSE is substantial, particularly in larger window sizes, which indicates that the proposed-TFT is better equipped to handle complex forecasting tasks. For example, in the largest window size (56,14), the Proposed-TFT model achieves improvements of up to 13.88% in MAE, 21.58% in MSE, and 9.36% in RMSE on the Pakistan SI dataset, and similar improvements are observed on the USA SI dataset. These improvements underscore the robustness of the Proposed-TFT model, making it a more reliable choice for accurate forecasting across different temporal scales and datasets.

The observed trends are likely due to the inherent differences in the datasets, as well as the ability of the models to generalize across varying conditions. The USA SI dataset might offer more consistent patterns or fewer outliers, allowing models like the Proposed-TFT to perform exceptionally well. In contrast, the Pakistan SI dataset might present more variability, which could explain the slightly higher error metrics across all models. Moreover, the inclusion of environmental factors makes the USA SI dataset more useful for getting higher performance as compared to the Pakistan SI dataset where those environmental parameters are not included.

V. CONCLUSION

This study highlights the importance of accurate solar irradiance forecasting for the efficient operation of solar power plants and the stability of power supply systems. We proposed a novel forecasting framework that integrates the Temporal Fusion Transformer (TFT) with Variational Mode Decomposition (VMD) to better capture the randomness and intermittency of solar irradiance. Our experimental results demonstrate that the VMD-TFT model significantly outperforms traditional models, including LSTM and ANN, achieving an R^2 score of 0.921, compared to 0.892 for Prophet. The interpretable outputs from our model provide valuable insights into the significance of both historical and future variables, aiding researchers and decision-makers in developing more accurate forecasting models. This work suggests future research directions such as expanding the dataset and exploring various time scales to enhance the reliability of solar irradiance predictions. The analysis of attention weight patterns may further uncover consistent temporal behaviors, such as seasonal fluctuations, thereby bolstering confidence in forecasting accuracy for both the scientific and industrial communities. Proposed method can enhance renewable energy management by improving the scheduling and integration of solar power into energy grids, ultimately reducing reliance on fossil fuels. Additionally, it can optimize smart grid operations, support agricultural

planning by aiding farmers in resource management, and improve building energy management by predicting solar energy availability. The model can also facilitate efficient electric vehicle charging by maximizing the use of renewable energy and contribute valuable data for climate studies and disaster management planning. These applications highlight the relevance and potential of our methodology in advancing sustainable practices across various sectors.

REFERENCES

[1] Y. Chu, Y. Wang, D. Yang, S. Chen, and M. Li, "A review of distributed solar forecasting with remote sensing and deep learning," *Renew. Sustain. Energy Rev.*, vol. 198, Jul. 2024, Art. no. 114391.

[2] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111799.

[3] L. Frías-Paredes, F. Mallor, M. Gastón-Romeo, and T. León, "Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors," *Energy Convers. Manage.*, vol. 142, pp. 533–546, Jun. 2017.

[4] M. J. Mayer, A. Szilágyi, and G. Gróf, "Environmental and economic multi-objective optimization of a household level hybrid renewable energy system by genetic algorithm," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 115058.

[5] M. Iqbal, "Solar radiation incident on tilted planes on the EARTHS surface," in *An Introduction to Solar Radiation*, M. Iqbal, Ed., New York, NY, USA: Academic, 1983, pp. 303–334.

[6] M. Taylor, P. Ralon, and A. Ilas, "Power to change: Solar and wind cost reduction potential to 2025," IRENA, Masdar City, United Arab Emirates, Tech. Rep., 2016.

[7] M. A. Hassan, L. Al-Ghussain, A. D. Ahmad, A. M. Abubaker, and A. Khalil, "Aggregated independent forecasters of half-hourly global horizontal irradiance," *Renew. Energy*, vol. 181, pp. 365–383, Jan. 2022.

[8] E. L. Delaney, J. M. McKinley, W. Megarry, C. Graham, P. G. Leahy, L. C. Bank, and R. Gentry, "An integrated geospatial approach for repurposing wind turbine blades," *Resour., Conservation Recycling*, vol. 170, Jul. 2021, Art. no. 105601.

[9] *Global Status Report 2019*, Renew. Energy Policy Netw. 21st Century (REN21), Paris, France, 2019, p. 336.

[10] L. S. Hoyos-Gómez, J. F. Ruiz-Muñoz, and B. J. Ruiz-Mendoza, "Short-term forecasting of global solar irradiance in tropical environments with incomplete data," *Appl. Energy*, vol. 307, Feb. 2022, Art. no. 118192.

[11] M. Martínez-Lavín, R. Villena-Ruiz, A. Honrubia-Escribano, J. C. Hernández, and E. Gómez-Lázaro, "Proposal for an aggregated solar PV power plant simulation model for grid code compliance," *Electr. Power Syst. Res.*, vol. 213, Dec. 2022, Art. no. 108676, doi: 10.1016/j.epsr.2022.108676.

[12] A. P. Farias-Rocha, K. M. K. Hassan, J. R. R. Malimata, G. A. Sánchez-Cubedo, and L. R. Rojas-Solórzano, "Solar photovoltaic policy review and economic analysis for on-grid residential installations in the Philippines," *J. Cleaner Prod.*, vol. 223, pp. 45–56, Jun. 2019, doi: 10.1016/j.jclepro.2019.03.085.

[13] A. Allouhi, S. Rehman, M. S. Buker, and Z. Said, "Up-to-date literature review on solar PV systems: Technology progress, market status and R&D," *J. Cleaner Prod.*, vol. 362, Aug. 2022, Art. no. 132339, doi: 10.1016/j.jclepro.2022.132339.

[14] Solargraf. (2022). *12 Ways to Facilitate a Solar Installation*. Accessed: Oct. 7, 2022. [Online]. Available: https://blog.solargraf.com/ways-facilitate-solarinstallation

[15] F. Ekinci, A. Yavuzdeğer, H. Nazlügül, B. Esenboğa, B. D. Mert, and T. Demirdelen, "Experimental investigation on solar PV panel dust cleaning with solution method," *Solar Energy*, vol. 237, pp. 1–10, May 2022, doi: 0.1016/J.SOLENER.2022.03.066.

[16] Z. Xia and J. A. Abu Qahouq, "Ageing characterization data of lithium-ion battery with highly deteriorated state and wide range of state-of-health," *Data Brief*, vol. 40, Feb. 2022, Art. no. 107727, doi: 10.1016/j.dib.2021.107727.

[17] H. Sharon, M. Vivar, and M. Fuentes, "A review on role of solar photovoltaic (PV) modules in enhancing sustainable water production capacity of solar distillation units," *J. Environ. Manage.*, vol. 320, Oct. 2022, Art. no. 115781, doi: 10.1016/j.jenvman.2022.115781.

[18] E. Costa, A. C. R. Teixeira, S. C. S. Costa, and F. L. Consoni, "Influence of public policies on the diffusion of wind and solar PV sources in Brazil and the possible effects of COVID-19," *Renew. Sustain. Energy Rev.*, vol. 162, Jul. 2022, Art. no. 112449, doi: 10.1016/j.rser.2022.112449.

[19] B. P. Mukhoty, V. Maurya, and S. K. Shukla, "Sequence to sequence deep learning models for solar irradiation forecasting," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–6.

[20] F. Al-Amri, F. Saeed, and M. A. Mujeebu, "Novel dual-function racking structure for passive cooling of solar PV panels–thermal performance analysis," *Renew. Energy*, vol. 198, pp. 100–113, Oct. 2022, doi: 10.1016/j.renene.2022.08.047.

[21] D. D. P. Rani, D. Suresh, P. R. Kapula, C. H. M. Akram, N. Hemalatha, and P. K. Soni, "IoT based smart solar energy monitoring systems," *Mater. Today, Proc.*, vol. 80, pp. 3540–3545, Jan. 2023, doi: 10.1016/j.matpr.2021.07.293.

[22] A. S. O. Ogunjuyigbe, T. R. Ayodele, and O. E. Oladimeji, "Management of loads in residential buildings installed with PV system under intermittent solar irradiation using mixed integer linear programming," *Energy Buildings*, vol. 130, pp. 253–271, Oct. 2016, doi: 10.1016/j.enbuild.2016.08.042.

[23] O. El Alani, M. Abraim, H. Ghennioui, A. Ghennioui, I. Ikenbi, and F.-E. Dahr, "Short term solar irradiance forecasting using sky images based on a hybrid CNN–MLP model," *Energy Rep.*, vol. 7, pp. 888–900, Nov. 2021, doi: 10.1016/j.egyr.2021.07.053.

[24] S. F. Rafique and Z. Jianhua, "Energy management system, generation and demand predictors: A review," *IET Gener., Transmiss. Distribution*, vol. 12, no. 3, pp. 519–530, Feb. 2018.

[25] H. Mukhtar, A. Afzal, S. Alahmari, and S. Yonbawi, "CCGN: Centralized collaborative graphical transformer multi-agent reinforcement learning for multi-intersection signal free-corridor," *Neural Netw.*, vol. 166, pp. 396–409, Sep. 2023.

[26] M. Xia, H. Shao, X. Ma, and C. W. de Silva, "A stacked GRU-RNN-based approach for predicting renewable energy and electricity load for smart grid operation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7050–7059, Oct. 2021.

[27] S. A. Nabavi, A. Aslani, M. A. Zaidan, M. Zandi, S. Mohammadi, and N. H. Motlagh, "Machine learning modeling for energy consumption of residential and commercial sectors," *Energies*, vol. 13, no. 19, p. 5171, Oct. 2020.

[28] Y. Yang, W. Li, T. A. Gulliver, and S. Li, "Bayesian deep learning-based probabilistic load forecasting in smart grids," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4703–4713, Jul. 2020.

[29] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2020.

[30] R. Yacef, M. Benghanem, and A. Mellit, "Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study," *Renew. Energy*, vol. 48, pp. 146–154, Dec. 2012.

[31] M. Q. Raza, N. Mithulananthan, and A. Summerfield, "Solar output power forecast using an ensemble framework with neural predictors and Bayesian adaptive combination," *Sol. Energy*, vol. 166, pp. 226–241, May 2018.

[32] D.-X. Niu, H.-F. Shi, and D. D. Wu, "Short-term load forecasting using Bayesian neural networks learned by hybrid Monte Carlo algorithm," *Appl. Soft Comput.*, vol. 12, no. 6, pp. 1822–1827, Jun. 2012.

[33] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[34] J. F. Torres, A. Troncoso, I. Koprinska, Z. Wang, and F. Martínez-Álvarez, "Big data solar power forecasting based on deep learning and multiple data sources," *Expert Syst.*, vol. 36, no. 4, p. 12394, Aug. 2019.

[35] S. Dhillon, C. Madhu, D. Kaur, and S. Singh, "A solar energy forecast model using neural networks: Application for prediction of power for wireless sensor networks in precision agriculture," *Wireless Pers. Commun.*, vol. 112, no. 4, pp. 2741–2760, Jun. 2020.

[36] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.

[37] N. Somu, G. R. M R, and K. Ramamritham, "A hybrid model for building energy consumption forecasting using long short term memory networks," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114131.

[38] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, "A novel CNN-GRU-Based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020.

[39] A. J. del Real, F. Dorado, and J. Durán, "Energy demand forecasting using deep learning: Applications for the French grid," *Energies*, vol. 13, no. 9, p. 2242, May 2020.

[40] R. Chen and M. Tao, "Data-driven prediction of general Hamiltonian dynamics via learning exactly-symplectic maps," in *Proc. ICML*, 2021, pp. 1717–1727.

[41] S. A. Haider, M. Sajid, H. Sajid, E. Uddin, and Y. Ayaz, "Deep learning and statistical methods for short- and long-term solar irradiance forecasting for Islamabad," *Renew. Energy*, vol. 198, pp. 51–60, Oct. 2022.

[42] F. Marchesoni-Acland and R. Alonso-Suárez, "Intra-day solar irradiation forecast using RLS filters and satellite images," *Renew. Energy*, vol. 161, pp. 1140–1154, Dec. 2020.

[43] D. C. Maddix, Y. Wang, and A. Smola, "Deep factors with Gaussian processes for forecasting," 2018, *arXiv:1812.00098*.

[44] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *Int. J. Forecasting*, vol. 36, no. 3, pp. 1181–1191, Jul. 2020.

[45] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 95–104.

[46] R. Sen, H.-F. Yu, and I. S. Dhillon, "Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting," in *Proc. NeurIPS*, 2019, pp. 4837–4846.

[47] S.-Y. Shih, F.-K. Sun, and H.-Y. Lee, "Temporal pattern attention for multivariate time series forecasting," *Mach. Learn.*, vol. 108, nos. 8–9, pp. 1421–1441, Sep. 2019.

[48] S. Wu, X. Xiao, Q. Ding, P. Zhao, Y. Wei, and J. Huang, "Adversarial sparse transformer for time series forecasting," in *Proc. NeurIPS*, 2020, pp. 17105–17115.

[49] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[50] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, I. Simon, C. Hawthorne, N. Shazeer, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck, "Music transformer," in *Proc. ICLR*, 2019, pp. 492–506.

[51] H. Mukhtar and M. U. G. Khan, "STMMOT: Advancing multi-object tracking through spatiotemporal memory networks and multi-scale attention pyramids," *Neural Netw.*, vol. 168, pp. 363–379, Nov. 2023.

[52] Y. Gao, Y. Ruan, C. Fang, and S. Yin, "Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data," *Energy Buildings*, vol. 223, Sep. 2020, Art. no. 110156.

[53] G.-F. Fan, X. Wei, Y.-T. Li, and W.-C. Hong, "Forecasting electricity consumption using a novel hybrid model," *Sustain. Cities Soc.*, vol. 61, Oct. 2020, Art. no. 102320.

[54] S. N. Mughal, Y. R. Sood, and R. K. Jarial, "A neural network-based time-series model for predicting global solar radiations," *IETE J. Res.*, vol. 69, no. 6, pp. 3418–3430, Aug. 2023.

[55] J. E. Hammond, R. A. Lara Orozco, M. Baldea, and B. A. Korgel, "Short-term solar irradiance forecasting under data transmission constraints," *Renew. Energy*, vol. 233, Oct. 2024, Art. no. 121058.

[56] L. S. Saoud, H. Al-Marzouqi, and R. Hussein, "Household energy consumption prediction using the stationary wavelet transform and transformers," *IEEE Access*, vol. 10, pp. 5171–5183, 2022.

[57] Q.-T. Phan, Y.-K. Wu, and Q.-D. Phan, "An approach using transformer-based model for short-term PV generation forecasting," in *Proc. 8th Int. Conf. Appl. Syst. Innov. (ICASI)*, Apr. 2022, pp. 17–20.

[58] J. Hu, J. Heng, J. Wen, and W. Zhao, "Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm," *Renew. Energy*, vol. 162, pp. 1208–1226, Dec. 2020.

[59] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Appl. Energy*, vol. 304, Dec. 2021, Art. no. 117766.

[60] V. Hoolohan, A. S. Tomlin, and T. Cockerill, "Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data," *Renew. Energy*, vol. 126, pp. 1043–1054, Oct. 2018.

[61] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renew. Sustain. Energy Rev.*, vol. 31, pp. 762–777, Mar. 2014.

[62] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021.

[63] H. Zhang, B. Li, S.-F. Su, W. Yang, and L. Xie, "A novel hybrid transformer-based framework for solar irradiance forecasting under incomplete data scenarios," *IEEE Trans. Ind. Informat.*, vol. 20, no. 6, pp. 8605–8615, Jun. 2024.

[64] Z. Wang, Y. Zhang, G. Li, J. Zhang, H. Zhou, and J. Wu, "A novel solar irradiance forecasting method based on multi-physical process of atmosphere optics and LSTM-BP model," *Renew. Energy*, vol. 226, May 2024, Art. no. 120367.

[65] Y. Ma, Q. Lv, R. Zhang, Y. Zhang, H. Zhu, and W. Yin, "Short-term photovoltaic power forecasting method based on irradiance correction and error forecasting," *Energy Rep.*, vol. 7, pp. 5495–5509, Nov. 2021, doi: 10.1016/j.egyr.2021.08.167.

[66] X. Wang, D. Pi, X. Zhang, H. Liu, and C. Guo, "Variational transformer-based anomaly detection approach for multivariate time series," *Measurement*, vol. 191, Mar. 2022, Art. no. 110791.

[67] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1748–1764, Oct. 2021.

[68] Aasim, S. N. Singh, and A. Mohapatra, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renew. Energy*, vol. 136, pp. 758–768, Jun. 2019.

[69] N. Khan, F. U. M. Ullah, I. U. Haq, S. U. Khan, M. Y. Lee, and S. W. Baik, "AB-Net: A novel deep learning assisted framework for renewable energy generation forecasting," *Mathematics*, vol. 9, no. 19, p. 2456, Oct. 2021.

[70] D. A. Widodo, N. Iksan, E. D. Udayanti, and Djuniadi, "Renewable energy power generation forecasting using deep learning method," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 700, no. 1, Mar. 2021, Art. no. 012026.

[71] D. Kaur, S. N. Islam, M. A. Mahmud, M. E. Haque, and A. Anwar, "A VAE-Bayesian deep learning scheme for solar generation forecasting based on dimensionality reduction," 2021, *arXiv:2103.12969*.

[72] 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.

[73] 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.

[74] J. Fang, L. Yang, X. Wen, H. Yu, W. Li, J. F. Adamowski, and R. Barzegar, "Ensemble learning using multivariate variational mode decomposition based on the transformer for multi-step-ahead streamflow forecasting," *J. Hydrol.*, vol. 636, Jun. 2024, Art. no. 131275.

[75] NREL. *What is the NSRDB?* Nat. Renew. Energy Lab. Accessed: Jul. 25, 2024. [Online]. Available: https://nsrdb.nrel.gov/about/what-is-the-nsrdb

[76] S. Pereira, P. Canhoto, and R. Salgado, "Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data," *Energy AI*, vol. 15, Jan. 2024, Art. no. 100314.

XINYANG HU is currently pursuing the bachelor's degree in environmental science with the College of Arts and Science, New York University of Environmental Science, with a focus on the intersection of technology and environmental phenomena. His work primarily focuses on the design and implementation of advanced sensor technologies for tracking and analyzing weather phenomena. He is particularly interested in exploring how these sensors can enhance their understanding of climate change and its impacts on global ecosystems. His research interests include diverse and interdisciplinary, encompassing the development and application of optical sensors in environmental monitoring.

• • •