HYBRID DEEP LEARNING APPLICATION FOR SOLAR IRRADIANCE FORECASTING MODEL

SAUMYA MISHRA[1], DEEPENDRA PANDEY[2], SAURABH BHARDWAJ[3]

[1]Research Scholar, Department of Electronics & Communication Engineering, Amity School of Engineering & Technology, Amity University, Lucknow Campus, Uttar Pradesh, India-226010
[2]Assistant Professor-III, Department of Electronics & Communication Engineering, Amity School of Engineering & Technology, Amity University, Lucknow Campus, Uttar Pradesh, India-226010
[3]Associate Professor, Department of Electrical & Instrumentation Engineering, Thapar Institute of Engineering & Technology, Thapar University, Patiala, Punjab, India-147001
E-mail: [1]msaumyabtech@gmail.com, [2]dpandey@lko.amity.edu, [3]saurabh.bhardwaj@thapar.edu

ABSTRACT

Secure scheduling, cost-effective power system operations, and the reduction of technical and financial risks for the electricity market all depend on accurate and trustworthy solar irradiance forecasting (SIF). This paper introduces a new hybrid, data-driven deep learning scheme called SIF. Various geographical and meteorological datasets are preprocessed, and then redundant variables are removed using feature selection (FS) and Grey wolf optimization (GWO-VMD)-enhanced variational mode decomposition. The refined data used for predicting the subsequence of solar irradiance by the BiLSTM-Attention network. Lastly, kernel density estimation for Gaussian kernel (KDE-Gaussian) is used to define prediction intervals for different confidence levels. The comparison shows that the FS-GWO-VMD-BiLSTM-Attention scheme performs better, reducing MAE/MAPE/MSE by 93%, 86%, and 98%, respectively to BPNN model. With narrower range, higher coverage, and deeper information of interval, the KDE-Gaussian approach produces outstanding coverage–width criterion (CWC) results that precisely reflect fluctuations in SI.

Keywords: *Solar Irradiance Forecasting, Bidirectional Long Short-Term Memory, Data-Driven Modeling, Interval Forecasting, Attention Mechanism*

1. INTRODUCTION

In the twenty-first century, renewable energy sources have gained enormous popularity, and sustainable technologies are constantly evolving on a global scale [1]. Among the various clean energy options, solar photovoltaic (PV) systems are regarded as one of the most widely used technologies. The International Renewable Energy Agency (IRENA) has projected that by 2055, the capacity of solar PV generation worldwide may reach 15,000 GW [2]. Solar power plants offer a rising share of electricity to grids, enhancing energy source diversification while also playing an important role in reducing environmental problems. However, large-scale PV integration presents substantial technological issues, such as preserving grid stability, assuring power quality, optimizing generation dispatch, and determining spinning reserve (SR) requirements [3, 4]. These issues originate from PV's changing working conditions, particularly the intermittent nature of the weather [5, 6].

1.1 Background

Higher power transients cause unpredictable power production, jeopardizing grid stability and impeding efficient electricity scheduling and dispatch [7, 8]. Historically, such fluctuation required increasing the system's spinning reserve (SR), which increased total operational expenses [9, 10]. As a result, reliable forecast of renewable energy generation is significant for precise SR planning, allowing operators to maximize reserve capacity, increase dispatch efficiency, and preserve grid dependability [11]. Beyond operational planning, power output forecasting facilitates the implementation of alternative mitigation strategies, such as the integration of energy storage systems [12, 13]. For PV installations, generation prediction can be accomplished primarily through two methods: (i) directly examining historical generation data or (ii) incorporating historical measurements of affecting factors—such as ambient temperature and solar irradiance—within an pricise physical PV model. The preliminary approach predicts future output by examining

historical time-series data of actual PV power generation. The second approach, on the other hand, starts by predicting the ambient temperature and solar irradiance across the intended horizon. These predicted values are then added to a physical PV system model. This model provides precise estimations of the plant's energy generation by relating temperature and irradiance to PV cell temperature and power output [14, 15]. Furthermore, this method takes into account a number of important elements that improve accuracy, including planned maintenance, cleaning procedures in sizable PV facilities, and specific PV cell characteristics like conversion efficiency and photon interaction behavior [16].

1.2 Motivation

Solar power generation is inherently unpredictable, intermittent and fluctuating, putting significant pressure on grid power balance. The expansion of solar energy installations in rapid mode has posed significant issues for power system planning, scheduling, operational stability, and the cost-effective use of solar resources. Short-term SIF is a critical tool for dispatch departments when developing generation schedules and changing operational tactics. Furthermore, precise forecasting results are required for solar power facilities to submit competitive bids and efficiently engage in the electricity spot market. As a result, precise and dependable SIF is crucial not only for maintaining safe dispatch and stable operation of power systems, as well as lowering balancing and backup costs, but also for mitigating the financial and technical risks that energy market industries confront.

1.3 Contributions and Novelties

This research introduces a novel hybrid forecasting framework named that combines advanced feature selection mechanism for deep learning models with the a modern optimization inspired by life style of gray wolves. Overall it has stage wise combination of feature selection, signal decomposition, deep learning, optimization, and interval forecasting in a single framework. The main novelty lies in combining GRA and PCC for feature selection, followed by GWO-VMD for generating reduced number of relevant features to forecast the irradiance data. The BiLSTM model is updated by using attention mechanism for capturing long-term dependencies and focus on significant temporal features. The use of Kernel Density Estimation (KDE-Gaussian) applied to produce the probabilistic forecasting intervals, making the proposed framework capable of both point wise and interval wise prediction of solar irradiance. This integrated structure significantly improves forecasting accuracy and reliability.

2. LITERATURE REVIEW

Several scholars have conducted comprehensive investigations on SIF. SIF forecasting time frames include ultra-short (<3 instants), short, mid-term, and long-term projections. Solar energy forecast typically focuses on either solar irradiance estimation or solar power output prediction. Output forecasts might be deterministic or probabilistic. Given the considerable uncertainty and unpredictability in solar power output, probabilistic forecasting, which includes the providing of prediction intervals (PIs), provides decision-makers with more meaningful and actionable information. The ARIMA model is regarded as a useful tool for fine-resolution, 24-hour solar radiation forecasting within conventional statistical techniques [17, 18]. On the other hand, sophisticated machine learning techniques like artificial neural networks (ANN) are found to be very accurate to predict both short-term (1–3 steps) and medium-term (12–24 steps) solar irradiance forecasts. With coefficients of determination ($R2$) of 0.96 in foggy conditions and 0.97 in bright conditions, ANN models demonstrated excellent performance [19]. For one-day SIF in grid-connected PV plants, another ANN-based approach demonstrated 94% accuracy on cloudy days and 98% accuracy on clear ones [20]. In addition to ARIMA and ANN, support vector machine (SVM) techniques have also demonstrated efficacy in PV output prediction. For example, Shi et al. [21] used SVM to classify days into four groups for model training: cloudy, foggy, sunny, and wet. However, SVM produced larger prediction errors for cloudy-day data, with an MRE of 12.42% versus 4.85% for sunny-day data. In regression forecasting, SVM beat ANN, with a R^2 value of 0.9562 compared to ANN's 0.9264 on a dataset with both gloomy and sunny days [22].

It has been demonstrated that LS-SVM produces precise estimates of solar irradiance with an $R2$ of 0.94 when applied to meteorological data [23]. With an $R2$ of 0.9929, Ekici [24] further proved its promise in one-day forecasting. Furthermore, SVM outperformed ANN in terms of prediction performance when combined with complementary techniques, raising $R2$ values from 0.8564 to 0.942 [25]. Notwithstanding these achievements, SVM and ANN have certain

drawbacks. While SVM and LS-SVM have poor interpretability and high computing costs, especially in probabilistic forecasting jobs, ANN requires huge datasets for training [26]. Hybrid models, which combine the advantages of several strategies, have been put out as practical answers to these problems. A hybrid RNN-ANN architecture, for instance, was presented by Castillo-Rojas et al. [27] and greatly enhanced short-term PV power generation projections.

Prediction accuracy has increased significantly by the combination of the ANN model of Multilayer Perceptron (MLP) and a RNN with LSTM units. Castillo-Rojas et al. [28] highlighted enhanced performance of this hybrid RNN–ANN method works with many input data. They used two variants of the model: one that relied exclusively on weather-related inputs, and the other that included both historical data of generated energy and meteorological factors. According to the finding of test results, the integrated-input model gave the best accuracy in predicting the weekly solar energy output.

For short-term multivariate prediction of input power, Xu et al. [29] presented an optimized structure based a sequential LSTM ensemble. It evaluated the performance of ensemble against a number of other approaches for wind and PV energy forecasting. In a similar vein, Li et al. [30] suggested a scenario-based forecasting system that created weekly PV power scenarios by combining a CNN and a Gated Recurrent Unit (GRU). Carrera et al. [31] created three different architectures: a hybrid system (PVHybNet) that included predictions and observations, an LSTM network for observations of weather conditions (obs-LSTM), and a Deep Feedforward Network (DFN) for weather forecasting (fcst-DFN). The hybrid framework produced higher accuracy, according to the results. For predicting global horizontal irradiance one hour in advance, Zang et al. [32] created a CNN–LSTM hybrid model, showing that the combined network outperformed CNN and LSTM when employed separately.

Due to heavy dependence on dynamic climatic parameters, forecasting of solar irradiance is still a challenging problem [33]. Numerous machine learning (ML) methods have been integrated with statistical models to enhance prediction accuracy, whereas others have incorporated numerical weather prediction (NWP) techniques. Nonetheless, NWP's performance degrades as the length of horizon for forecasting becomes larger [34]. Their unreliability and interpretability in cloudy conditions, where solar irradiance is highly variable, is a long-standing problem [35–37]. To address these problems, Homouz and Maalouf [38] introduced the Truncated-Regularized Kernel Ridge Regression (TR–KRR) approach, which minimized computational complexity significantly and improved interpretability while maintaining accuracy equal to or superior to Support Vector Machines (SVM). Moreover, TR–KRR has also been found to be usable in a wide range of applications [39–44].

A new hybrid framework for predicting sun irradiance was created and verified using data gathered from Medford and Seattle (USA). The model, which is designed for medium - to short - term time horizons, strikes a balance between high prediction accuracy and computing economy to solve the ongoing problem of forecasting under sporadic overcast situations. The strategy lowers operating and reserve costs by facilitating more efficient power generating scheduling. The increasing need for precise, accurate, and effective PV power forecasting methods is satisfied by this study. The hybrid design creates a scalable and adaptable solution by combining deep learning, optimization techniques, and Variational Mode Decomposition (VMD). The suggested technique's improved performance was validated by comparison with persistence models and traditional forecasting methods; the results demonstrated that it is a reliable and extremely accurate forecasting tool.

The modern era of information technologies based on artificial intelligence, cloud computing, and big data analysis have become increasingly popular with the onset of the 5G era. Solar farms today utilize Supervisory Control and Data Acquisition (SCADA), Geographic Information System (GIS), and NWP to generate and save tremendous amounts of datasets. The multi-source, multi-dimensional, and multi-modal character of these datasets significantly increases the complexity of data processing. Solar irradiance forecasting (SIF) has therefore been a field of research focus owing to the integration of advanced deep learning (DL) techniques with data-driven approaches [33, 42]. Data space infrastructure and storage services, as shown in Figure 1, facilitate solar power plants to diligently retain valuable information and systematically handle diverse data streams, improving effective operation and informed management decisions.

Most SIF studies are model-based with insufficient focus on the collection, management, storage, and systematic analysis of historical power

and meteorological data. Latent information and coupling mechanisms embedded in solar-related data are also given insufficient attention. Traditional approaches typically only support deterministic point prediction, don't provide detailed error quantification and uncertainty estimates, and hence are less effective in real applications. This article proposes a data-driven platform that pre-processes and processes multi-source, multi-dimensional data collected from real records by applying a hybrid deep learning algorithm to bypass these limitations. A framework developed specifically for ultra-short-term SIF that enhances the model's prediction accuracy and generalization ability.

2.1 Knowledge Gap Addressed

The literature review on latest forecasting schemes shows that most of the state of art methods focuses on deterministic forecasting and ignores the uncertainty, error quantification, and interval reliability. Most previous works lacks the stage of robust data preprocessing and feature optimization. Additionally, traditional methods that are in conventional use neglect the multi-dimensional aspects of the meteorological data and do not included probabilistic approach for finding the confidence intervals for scheduling and risk assessment. The proposed research work bridges this gap by the hybrid mechanism that optimizes input features, decomposes raw data, performs learning, and quantifies prediction uncertainty with accurate confidence intervals.

Traditional statistical models like ARIMA give acceptable perform for short-term forecasting but fails in situation of rapid change in weather conditions, making it unreliable for climates with dynamic response [17, 18]. ML models such as ANN has high accuracy under clear conditions (R^2 = 0.96–0.97) but its performance deteriorate in cloudy or foggy weather and require large datasets for training the model [19, 20]. SVM approaches improve accuracy but still show high error during cloudy days that is observed in terms of MSE of 12.42% compared to 4.85% for sunny conditions [21]. Additionally it faces issues of high computational complexity and lack interpretability for probabilistic forecasting applications [23, 26]. The literature reviewed in this article focus mostly on deterministic forecasting and does not offer prediction intervals or uncertainty estimation. Hybrid deep learning models like CNN-LSTM [32], hybrid RNN-ANN [27, 28], and GRU-CNN [30] have improved forecasting accuracy but still do not include interval forecasting or confidence-

bound prediction. Many literatures do not apply data decomposition techniques, causing errors when dealing with noisy and non-stationary irradiance signals. Significant drop in prediction accuracy observed if no decomposition or noise reduction is applied [33, 34]. Some of the works shows high variability under cloudy weather, poor feature selection, and complex multi-source data and giving unstable results and poor generalization in forecasting [35, 37]. In this way, a complete forecasting mechanism that is including data preprocessing, feature selection, signal decomposition, deep learning, and uncertainty estimation is missing in current studies.

2.2 Need behind the Proposed Work

Solar irradiance has high level of unpredictability due to random variations in atmospheric conditions. This ambiguity creates critical challenges in grid stability, reserved power source estimation, and making strategies for market bidding. Traditional methods face difficulty in forecasting due to signal fluctuations, high noise, and uncertainty that consequently result in financial and operational losses. In this way the solar irradiance forecasting needs an efficient intelligent model to deliver accurate forecasting along with confidence intervals. This work fulfills this need by integration advanced feature, signal decomposition, and probabilistic modeling that helps in better operational planning, cost optimization, and risk mitigation for solar power plants and grid operators.

The abrupt fluctuations in solar irradiance impact the grid operations and force to maintain large energy storage reserves. It increases the operational costs [7, 9, 10]. It has been observed from literatures that battery storage size, optimal grid scheduling, and bidding strategies in power sector depend heavily on short-term forecasts [12, 13]. It is also found that weather-dependent fluctuations and lack of error estimation lead to poor operational decisions and financial losses [14, 16]. Models like CNN-LSTM [32] and PVHybNet [31] improve accuracy but these works do not address uncertainty quantification or interval prediction. This creates the need for a forecasting system that not only predicts accurate values but also provides prediction intervals. Literature supports that combining decomposition, feature selection, and deep learning improves forecasting reliability and supports energy market under uncertain conditions [34, 35, 38].

In addition, a non-parametric estimation method was employed to estimate solar irradiance and give

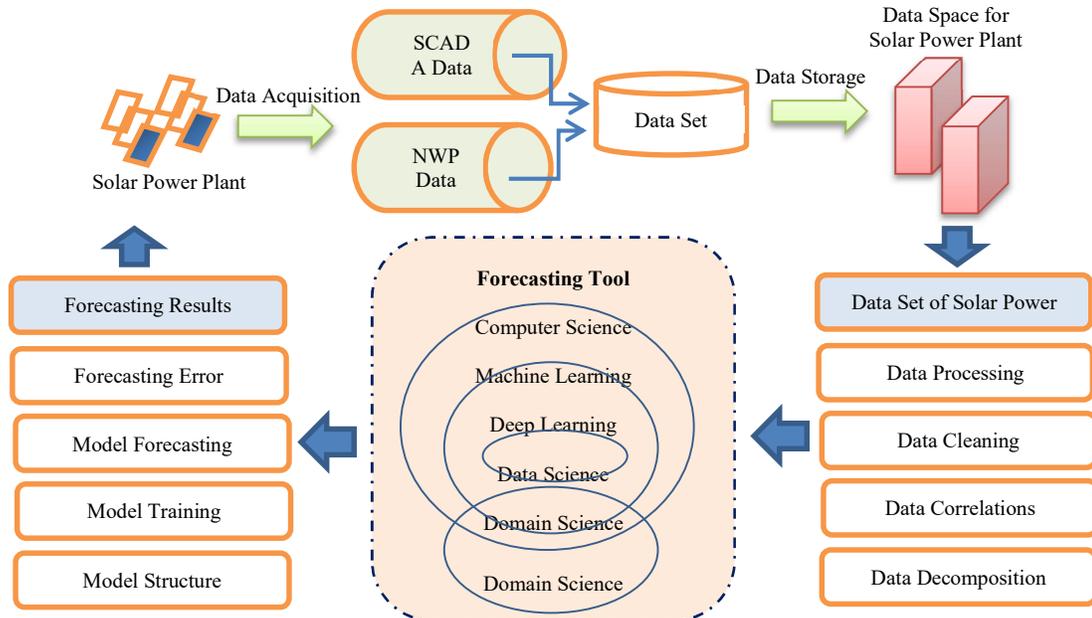more information in supporting the decision-making process.

Figure 1: Solar irradiance forecasting based on data-driven

The overview of the main contributions and innovations is defined as:

(1) The data-driven models performance is heavily reliant on the quality of the input data and the relation in the output and input variables. This is tackled by preprocessing the raw data and using Grey Relational Analysis (GRA) and Pearson Correlation Coefficient (PCC) in analyzing correlations of input features and their association. This eliminates redundant features and selects variables that are highly influential to the model. Accuracy and complexity can be balanced against each other due to feature selection (FS).

(2) The very high variability and randomness of SI time series severely weaken predicting stability and accuracy. The data were decomposed and stabilized by VMD and optimization achieved through Grey Wolf Optimizer (GWO) to reduce this. GWO determines the quadratic penalty coefficient and optimal number of decomposed modes for VMD through envelope entropy minimization as the fitness function.

(3) To leverage the ability of the LSTM to learn temporal features, this research utilizes a Bidirectional LSTM (BiLSTM) that captures temporal relationships between future and past time steps in a series. The dual-directional processing works for enhancing the capability of the model to deal with complex time-series forecasting and enables it to learn deeper latent features. BiLSTM enhances accuracy by recognizing patterns between output value of solar irradiance and input features and then further enhancing the SI representation at different frequencies.

(4) An attention mechanism is built into the model to counter a common shortcoming of BiLSTM models when dealing with very long time-series. By assigning sequence elements probabilistic weights, the attention mechanism effectively brings out key features while reducing the likelihood of information loss. Accuracy and computational efficiency are enhanced with integration. The reliability and strength of the hybrid framework, designated as FS-GWO-VMD-BiLSTM-Attention, were verified in comparison to standard forecasting methods through numerous frameworks.

(5) Kernel Density Estimation (KDE) is employed to assess errors in point forecasting and establish lower and upper bounds for SIF prediction curve to address the uncertainty inherent in SI forecasting. Prediction intervals (PIs) are generated with different confidence levels using the KDE-Gaussian method, which provides a probability perspective beyond deterministic predictions.

The effectiveness and reliability of the proposed interval forecasting system were also demonstrated by comparison with alternative methods using standard interval performance measures. This is the

layout details of the rest of the article. The proposed methodology involving feature selection, data decomposition, and point and interval SIF approaches is presented in Section 2. The comparative results of point and interval projections of different models are covered in Section 3. Finally, the summary of major findings and contributions of this research are given in Section 4.

3. MEHODOLOGY

Figure 3 shows the proposed ultra-short-term SIF framework, which is based on DL techniques and the data-driven paradigm. Feature selection (FS) is the next step after thorough data pretreatment by Famework. Decomposed using VMD optimized by GWO, training data was then prepared.BiLSTM and an attention mechanism are combined in this hybrid DL model to provide precise and computationally effective SIF. Prediction intervals (PIs) within multiple confidence levels calculated by interval forecasting, which is done using KDE applied to point forecasting errors, as a complement to point predictions.
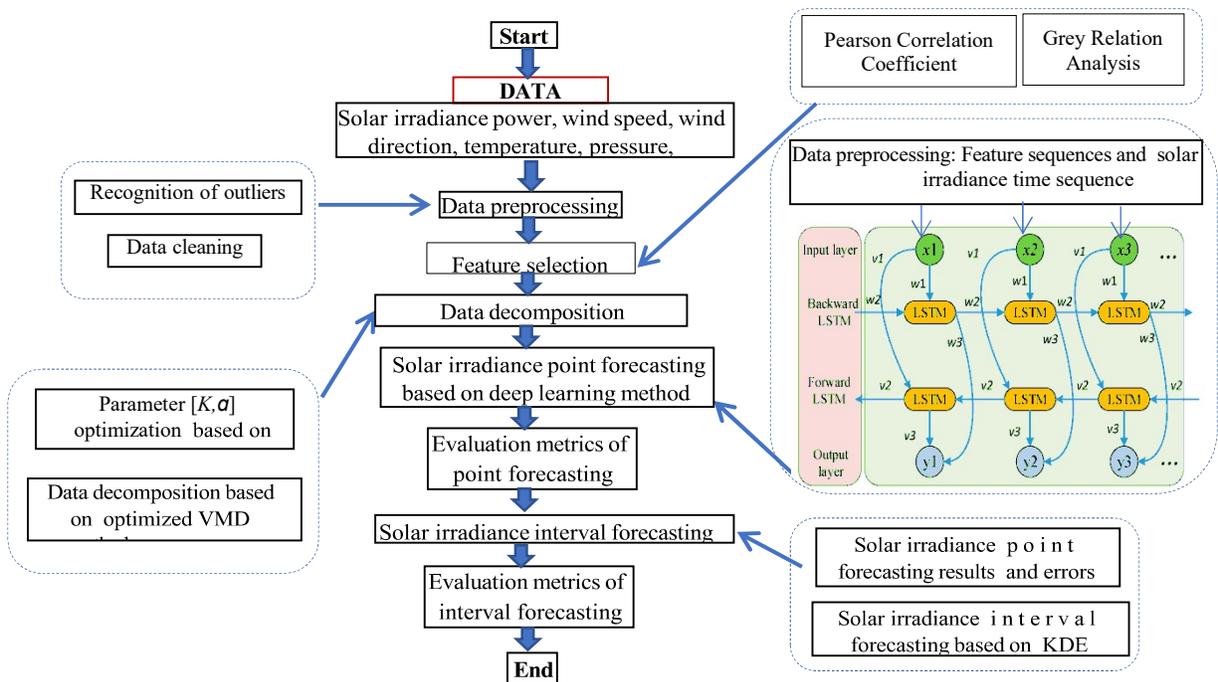

Figure 2: Systematic diagram for block wise representation to explain proposed methodology

3.1 Feature Selection Scheme
3.1.1 Pearson correlation coefficient (PCC)
A traditional statistics based technique for measuring a linear relation in two variables PCC. Equation (1) presents the computational formula. The most useful variables for predicting are chosen based on PCC values, which range from -1 to 1. Higher absolute values signify stronger correlations.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_y)]}{\sqrt{D_x}\sqrt{D_y}} \qquad (1)$$

where, $\rho_{X,Y}$ indicates the coefficient of correlation in variable X and Y, cov(X,Y) denotes the covariance of variable X with respect variable Y, σ_X and σ_Y denotes the standard deviation, respectively. μ_X and μ_Y indicate the mean values, DX and DY are variances of X and Y. (x_i, y_i) ($i = 1, 2,…, n$) represents the set as the sample from total (X,Y). The calculation method of sample PCC $r_{x,y}$ is shown in Eq. (2).

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (2)$$

$\bar{x} = {1}/{n}\sum_{i}^{n} x_i$ and $\bar{y} = {1}/{n}\sum_{i}^{n} y_i$ represent mean values of sample x_i and y_i respectively.

3.1.2 Grey relation analysis

Grey Relational Analysis (GRA) determines how close the variables are to one another by examining the geometric similarity of the data sequence inside

a common reference coordinate system. Its basic idea is to use sun irradiance as the output benchmark and use the Grey Relational Degree (GRD) to measure the relationship between solar irradiance data and influencing factor sequences.

The method's ease of computation and strong practical applicability are praised; the precise calculation procedure is included in the Supplementary Materials.
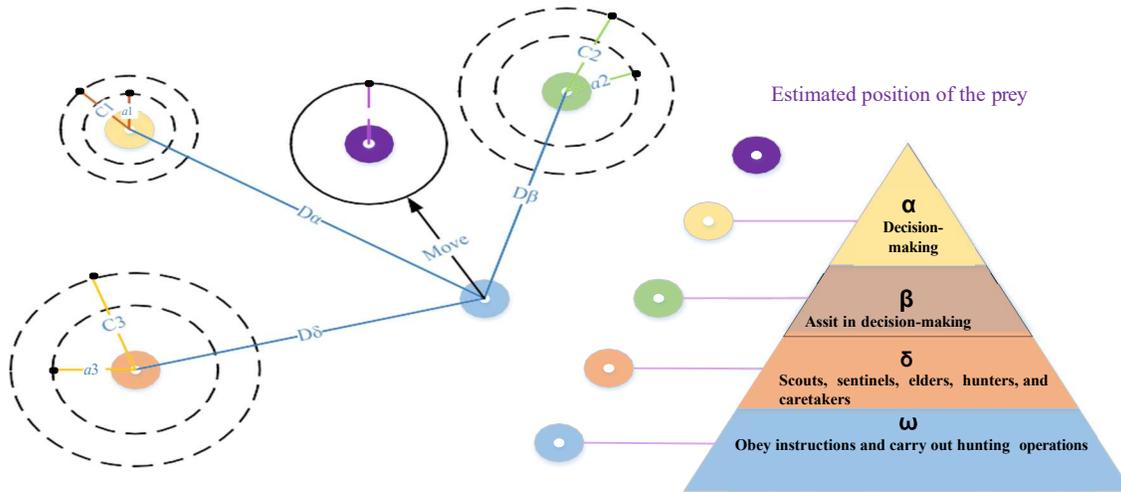

Figure 3: Diagram of GWO algorithm

γ_{0i} denotes the GRD between feature index sequence X_i and solar irradiance data sequence X_0, and the greater the γ_{0i} is, the greater the correlation degree in the feature variables as input and solar irradiance. Therefore, a threshold γ^* can be determined, and when $\gamma_{0i} \geq \gamma^*$, it is considered that the feature index has a greater impact on solar irradiance. When $\gamma_{0i} < \gamma^*$, this index is eliminated. The value of γ^* can be determined according to the resolution coefficient ξ, or according to the calculation results of GRD combined with practical problems.

In order to determine how closely two variables are related, Grey Relational Analysis (GRA) compares the geometric patterns of their data sequences using a common reference coordinate system. The Grey Relational Degree (GRD), which measures the correlation between sequences of input feature data and the target solar irradiance data sequence, is the fundamental idea behind GRA. The simplicity, computational effectiveness, and operational flexibility of this approach make it beneficial. The Supplementary Materials include step-by-step calculations. In the GRA framework, γ_{0i} denotes the GRD between the feature sequence X_i and the solar irradiance data sequence X_0. Higher values of γ_{0i} indicate stronger correlations between the target and the feature variable. A threshold γ^* is established such that features with $\gamma_{0i} \geq \gamma^*$ are retained as influential inputs, whereas those with γ_{0i} $< \gamma^*$ are discarded. The threshold γ^* can be determined using the resolution coefficient ξ or based on the GRD outcomes in combination with practical considerations, ensuring that only relevant features are selected for forecasting.

3.2 Meteorological Data Decomposition method

3.2.1 Variational mode decomposition

Variational Mode Decomposition (VMD) is a fully adaptive, non-recursive signal processing technique [45]. The first step in the process is to determine how many modes are required to decompose a non-stationary signal sequence. By adaptively estimating ideal center frequencies and finite bandwidths, VMD effectively separates the signal's frequency spectrum and extracts intrinsic mode functions (IMFs). The method solves the related variational optimization problem and yields optimal modal components. Unlike earlier techniques, VMD, which has a rigorous mathematical basis, simultaneously addresses endpoint effects and mode aliasing issues associated with Empirical Mode Decomposition (EMD) and mitigates residual noise issues observed in Ensemble EMD (EEMD).The non-stationary solar irradiance quantity in this study is broken down into modal components (t), where $k=1,2,...,K$, in order to prepare the data for hybrid DL forecasting.

3.2.2 Grey wolf optimization

GWO is a meta-heuristic program that follows intelligence under the behavior of swarm [46] that adaptively modifies information flow and convergence variables to preserve equilibrium between local exploitation and global exploration by imitating the tactics of hunting and social hierarchy of grey wolves. As a result, GWO performs well on optimization problems, obtaining fast convergence and great accuracy.

As seen in Fig. 3, grey wolves form a rigid pyramidal social structure in which members take on various duties based on their position. The grey wolf population in search space is initialized at random at the start of the method. Wolves at top-ranking are α, β, and δ estimate the positions of prey, while subordinate wolves (ω) update their positions based on relative distances to leaders. Wolves surround and converge on prey through repeated updates, which matches the best possible solution in the search space.

3.2.3 GWO-VMD based data decomposition stage

In order to reduce non-stationarity and make it appropriate for applications like solar irradiance research, VMD efficiently breaks down extremely complicated and nonlinear sequences in multiple subsequences at different frequencies. The quadratic penalty coefficient (α) and total number of modes (K) impact significantly on the quality of components that are produced during decomposition.

The GWO-VMD framework uses process of minimization for envelope entropy as the fitness function to maximize the K and α for reducing the sensitivity of VMD towards the variation in parameters. The original solar irradiance sequence's sparsity is measured by envelope entropy; a greater entropy means the intrinsic mode function (IMF) possess higher noise percentage and less feature information, whereas lower entropy suggests cleaner and more informative components. The entropy signal envelope E_p for sample k=1, 2, ..., N, is computed according to Eq. (3), guaranteeing that VMD generates the most significant and reliable components.

$$\begin{cases} E_p = -\sum_{j=1}^{N} p_k \log (p_k) \\ p_k = a(k) \Big/ \sum_{j=1}^{N} a(k) \end{cases} \quad (3)$$

p_k is normalized value of $a(k)$, $a(k)$: envelope signal on applying the Hilbert demodulation of the modal component $x(k)$ after VMD.

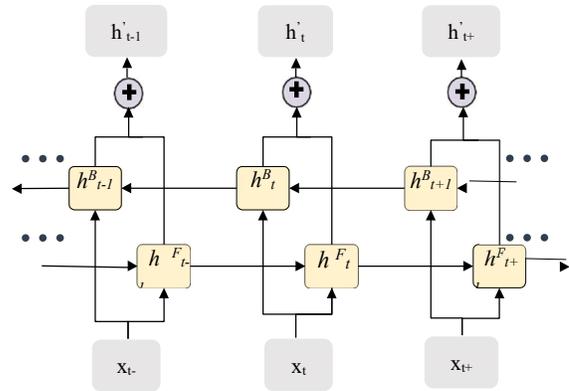### 3.2.4 Deep learning method based point forecasting

Figure 4: Diagram showing LSTM circulation unit

Bidirectional long short-term memory

A customized RNN that solves the disappearing and problems exploding gradient in the conventional RNNs is the LSTM network [47]. Bidirectional LSTM (BiLSTM), one of the most popular LSTM versions, was introduced to further improve LSTM's capabilities.

Two LSTM cells that process the identical input sequence in opposing temporal directions make up each hidden layer unit in BiLSTM. In order to record earlier temporal dependencies, the forward LSTM cell propagates forward after receiving outputs from the previous time step, see Fig. 4. In order to include future temporal information, the backward LSTM cell propagates the outputs from the subsequent time step backward. Bidirectional structure improves the ability of SIF models to learn and increase their accuracy while also improving the extraction of temporal characteristics from time series. Eqs. (4)–(6) illustrate how the BiLSTM unit's total output h' is calculated.

$$h_t^F = \vec{h}_t = LSTM(h_{t-1}, x_t, c_{t-1}) \quad (4)$$

$$h_t^B = \vec{h}_t = LSTM(h_{t+1}, x_t, c_{t+1}) \quad (5)$$

$$h_t' = h_t^F \oplus h_t^B \quad (6)$$

h_t^F & h_t^B represents values of output calculated at the forward & backward LSTM unit, here $\oplus$ represents vector addition operator.
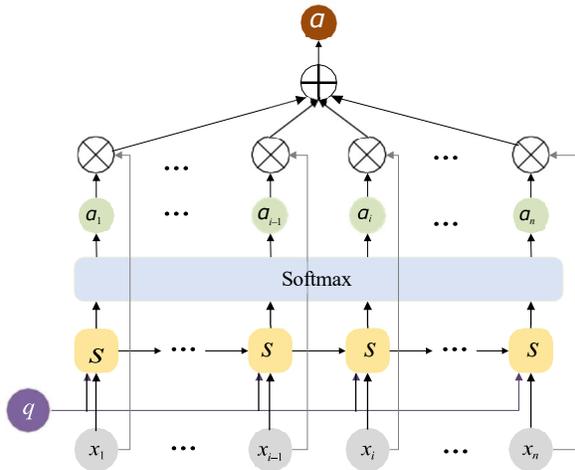
Figure 5 : Soft attention mechanism architecture

Attention mechanism

A complex cognitive mechanism called attention allows one to ignore less significant information in favor of concentrating on pertinent inputs at any given time. Motivated by this, ANN focuses processing power on important data, identifies key elements in a sequence, and improves SIF's efficiency and prediction accuracy [48].

Soft attention focuses the model's attention on information that is related to the concerned task by using trainable weights to assign priority to global input variables. BiLSTM incorporates soft attention to improve accuracy by addressing the issue of information loss in long-term series by employing probabilistic weights instead of random weights. The input sequence is represented by $X=[x_1,x_2,...,x_i,...,x_n]$ in Figure 5, the query vector is represented by q, and the attention score function that assesses the association between input and query is represented by $S(x_i,q)$. First, the input element X's attention distribution α_i was computed. Second, weighted average of input features computed according to attention distribution α, emphasizing contribution of critical features in forecasting SI.

3.2.5 Solar irradiance point forecasting method based on BiLSTM-Attention

The four layers of the BiLSTM-Attention-based point forecasting model are input, feature extraction, feature correlation, and output (Fig. 2). Outlier identification, missing value imputation, cleaning of data, selection of feature, and decomposition SI data are all part of the data preprocessing step, which guarantees high-quality model input (fig.7). The heart of the model is made up of feature correlation and extraction layers,

which allow for thorough feature mining. In the first stage, feature correlation—which uses BiLSTM units—captures temporal dependencies and creates correlations between SI time series and input feature sequences. Different input feature contributions to SI are identified by the feature extraction layer and attention mechanism. It accentuates the significance of crucial temporal information, completely captures long-term temporal interdependence, and highlights elements that have a major impact on SI. BiLSTM-Attention point forecasting efficiently mines temporal and feature information by combining feature processing phases, yielding precise results.

Method of interval forecasting using kernel density estimation (KDE)

Predicting the lower and upper bounds of SIF and determining related prediction interval (PI) for specified confidence level are the main goals of SI interval forecasting. Reducing operational hazards in power systems and giving decision-makers more thorough information are two benefits of having a clearly defined PI. In this study, the KDE approach is applied to quantify and quantitatively examine the point forecasting errors of SIF. Combining interval estimates of point forecasting errors with point forecasting outcomes yielded lower and upper bounds of the SI interval forecasts for different confidence levels. KDE is a non-parametric, data-driven model used for estimation of the probability density. It provides the ease of use, computational effectiveness, and is especially well-suited for predicting densities in situations when the distribution of the underlying data is uncertain [49]. This improves robustness and information content by facilitating the creation of trustworthy interval forecasts to supplement point predictions.

$e = [e_1, e_2 ,..., e_n]$:forecast errors of solar irradiance, and the KDE of the total density function $\hat{f}(e)$ at point e defined in Eq. (7).

$$\hat{f}_h(e) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{e - e_i}{h}\right) \quad (7)$$

The bandwidth parameter is represented by h, the kernel function is represented by $(\cdot)$, & total number of samples indicated as N in the KDE framework. There are various kinds of kernels, such as Quartic, Epanechnikov, Triangular, Gaussian, and Uniform. The most often used of these is the Gaussian kernel, whose mathematical form may be found in Eq. (8). As stated in Eq. (9) [50], the Epanechnikov kernel, on the other hand, is frequently regarded as optimal in the sense of

minimizing mean squared error (MSE) with just a slight decrease in computing efficiency.

$$K(\eta) = 1/\sqrt{2\pi}\, e^{\left(-\eta^2/2\right)} \qquad (8)$$

$$K(\eta) = \begin{cases} \dfrac{3}{4}[1-\eta^2], & \eta \in [-1,1] \\ 0, & else \end{cases} \qquad (9)$$

$\eta = (e-e_i) / h$ for specified significance level α, probability for predicting SIF value falls under the PI range is defined as prediction interval nominal confidence (PINC).

$PINC = (1-\alpha).\, 100\,\%$
for significance level α, PI $\hat{I}_i^{\alpha}$ for ith sample is given by:

$$\hat{I}_i^{\alpha} = \hat{U}_i^{\alpha} - \hat{L}_i^{\alpha} \qquad (11)$$

here, U and L respectively refers the lower and upper boundaries of PI.

The input variables selection is based on the physical relevance and statistical correlation with solar irradiance. Geographical factors (latitude, longitude, altitude) are included to consider the spatial variation, while temporal dependency is incorporated in terms of month of the year. The meteorological parameters are temperature, humidity, wind speed, sunshine duration, and solar radiation in ast month. These are chosen due to their direct influence on irradiance intensity and atmospheric transmission. PCC and GRA employed to identify the most influential predictors. Results (Fig. 9 and Fig. 10) indicate the strong correlations for sunshine duration, maximum and minimum temperatures, while altitude and wind speed showed weaker association and hence excluded.

4. RESULTS AND DISCUSSION

4.1 Data Preprocessing

Data was gathered from Uttar Pradesh, India, to confirm the efficacy of the suggested point and interval short-term SIF model. With 250 million people spread throughout 75 districts, it is the most densely inhabited state in the nation. 14 stations' worth of meteorological and SI data were retrieved of the UN Food and Agriculture Organization (FAO) using the CROPWAT 8.0 and CLIMWAT 2.0 tools. Geographical information is taken into account, including latitude (LAT), longitude (LON), and altitude (ALT). Minimum temperature (MIT), maximum temperature (MAT), wind speed (WIN), humidity (HUM), sunshine duration (SUN), and solar radiation (RAD) are all averaged

annually in CSV data. SI output variable $Y_N \times 1$ and input variables ALT, LAT, LON, MOY, MIT, HUM, and WIN as matrix $X_N \times 7$ with $N=2000$ samples. SI's time series (Fig. 8) shows significant volatility. For providing a robust basis for training and validation, the SI dataset additionally incorporates a number of meteorological parameters.
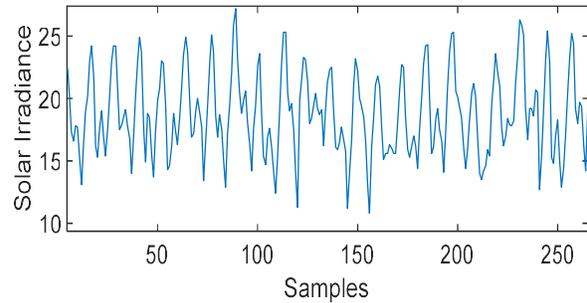

Figure 6: Raw Solar Irradiance time series data plot
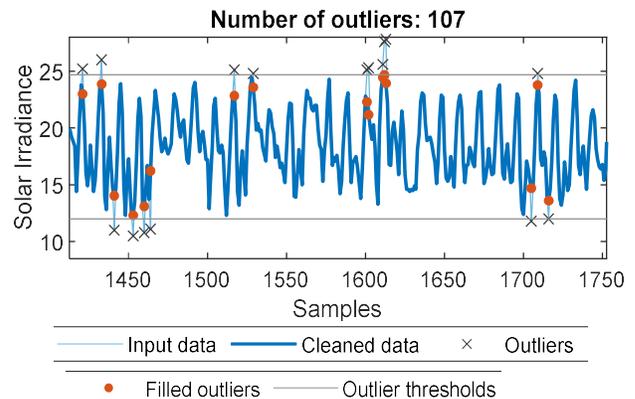

Figure 7: Removing outlier. Cleaning method: Modified Akima Cubic Interpolation, Detection method: quartiles, Threshold value: 0.85

A statistical analysis of the dataset revealed the existence of outliers and missing values, which reduces forecasting model accuracy and causes prediction mistakes. The dataset was preprocessed before the model was created. Outlier identification is done first, records with missing entries are eliminated, and any remaining anomalies are corrected. The dataset has a monthly temporal resolution, and missing values are imputed using the average of nearby observations. The second step is data cleaning. Third, feature preprocessing done, irradiance direction converted using trigonometric functions, and sine of incidence angle substituted original values. A refined dataset of 2000 samples was obtained after completion. 70% data was kept for training of the model, 20% for interval forecasting error estimation and hyperparameter optimization, and

the remaining 10% was used as a test set. Robust model creation, tuning, and final performance validation were guaranteed by this section.
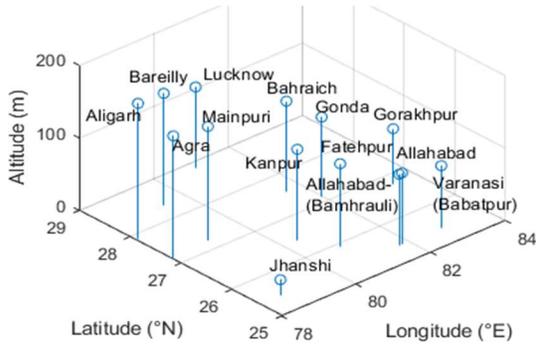

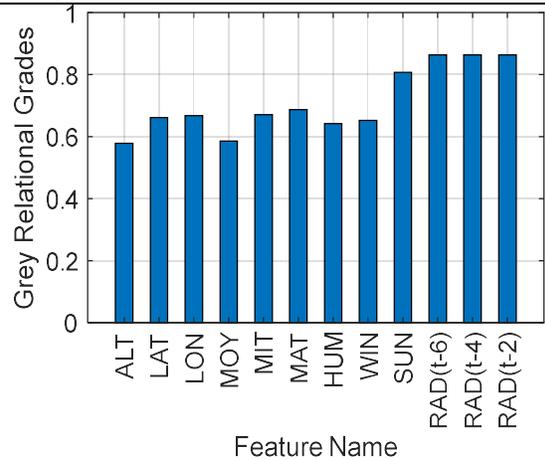*Figure 8: Geographical details of location used to collect data*

Figure10: GRG between solar irradiance and different features

4.2 PCC-GRA application for feature selection

For reducing the complexity of modeling process of the data-driven SIF, the GRA and PCC method are applied for analysis of influencing factors on solar irradiance. The calculated results are showing in the Fig. 9 and 10. The PCC between solar irradiance at different delay is strong, but the linear correlation between MOY and LAT is weak. Solar irradiance is showing a strong positive correlation with SUN, MIT and MAT.
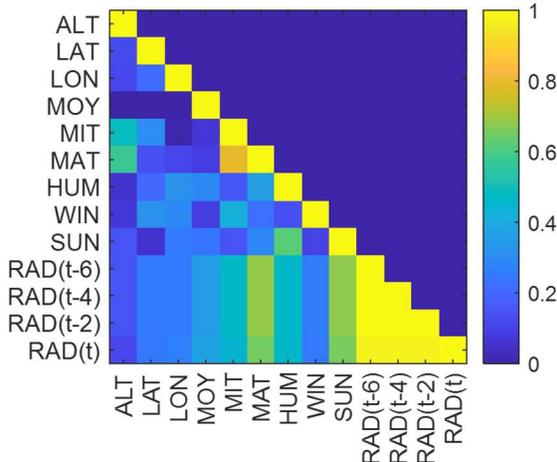

Figure.9: PCC between solar irradiance and different features

4.2.1 GWO-VMD application for solar irradiance data decomposition

Results outcomes are revealing that after data preprocessing, the PCC in RAD(t) and RAD at t-2,t-4 and t-6 at different delay has been improved significantly, but the PCC value in between solar irradiance and MOY and ALT has been decreased. Moreover, the GRD between solar irradiance and MAT/SUN is improved significantly, but the GRD between ALT and WIN is low always, hence this feature is eliminated. On the basis of results analysis of GRD and PCC, with other variables are taken as the feature variables as input of the SIF model.

The raw solar irradiance sequence exhibits significant changes, as seen in Figure 11 (Top), which is indicative of interference between several frequency components. The challenge of producing precise forecasts is complicated by this unpredictability. In order to improve forecast accuracy and stability, the GWO-VMD method is used. The number of decomposed modes and the quadratic penalty factor are two crucial VMD parameters whose ideal values are found using the GWO algorithm in this architecture. With a starting population of 20 potential solutions, the optimization is run for a maximum of 30 iterations. The optimization issue is two-dimensional since there are only two variables involved. Consequently the volatile irradiance data is broken down into a number of smoother sub-series with unique frequency characteristics.
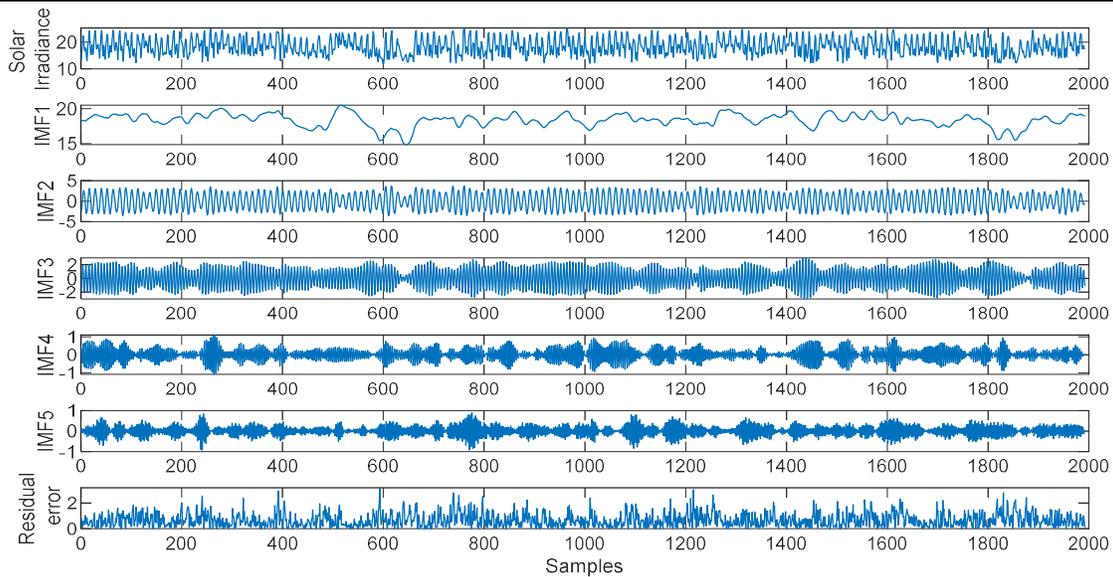
Figure 11: Decomposition results of solar irradiance data using GWO-VMD

Table 1: Comparative results summarized for demonstrating the models performance for different types of point forecasting

Model	RMSE (MW)	MAE (MW)	MSE (MW)	MAPE	R^2
FS-GWO-VMD-BiLSTM- Attention	0.76	0.48	0.57	3.1%	0.9997
GWO-VMD-BiLSTM-Attention	1.3	0.68	1.7	3.2%	0.9986
FS-BiLSTM-Attention	1.5	0.86	2.1	3.7%	0.9983
BiLSTM-Attention	1.4	0.87	2.2	3.9%	0.9982
FS-GWO-VMD-BiLSTM	1.5	0.96	2.2	4.4%	0.9984
GWO-VMD-BiLSTM	1.6	1.04	2.3	4.8%	0.9983
FS-BiLSTM	2.1	1.12	4.4	4.9%	0.9962
BiLSTM	2.2	1.19	4.5	5.1%	0.9961
FS-GWO-VMD-LSTM	2.4	1.3	5.7	5.3%	0.9953
GWO-VMD-LSTM	2.4	1.5	5.8	5.6%	0.9952
FS-LSTM	2.6	1.6	6.6	6.9%	0.9947
LSTM	2.7	1.7	7.3	7.4%	0.9941
FS-GWO-VMD-BPNN	3.5	2.3	12.3	12.7%	0.9901
GWO-VMD-BPNN	5.1	3.9	25.5	14.1%	0.9795
FS-BPNN	6.4	4.8	40.5	18.2%	0.9674
BPNN	10.1	8.1	116.7	21.6%	0.9063

Results outcomes are revealing that after data preprocessing, the PCC in RAD(t) and RAD at t-2,t-4 and t-6 at different delay has been improved significantly, but the PCC value in between solar irradiance and MOY and ALT has been decreased. Moreover, the GRD between solar irradiance and MAT/SUN is improved significantly, but the GRD between ALT and WIN is low always, hence this feature is eliminated. On the basis of results analysis of GRD and PCC, with other variables are taken as the feature variables as input of the SIF model.

The raw solar irradiance sequence exhibits significant changes, as seen in Figure 11 (Top), which is indicative of interference between several frequency components. The challenge of producing precise forecasts is complicated by this unpredictability. In order to improve forecast accuracy and stability, the GWO-VMD method is used. The number of decomposed modes and the quadratic penalty factor are two crucial VMD parameters whose ideal values are found using the GWO algorithm in this architecture. With a starting population of 20 potential solutions, the optimization is run for

a maximum of 30 iterations. The optimization issue is two-dimensional since there are only two variables involved. Consequently the volatile irradiance data is broken down into a number of smoother sub-series with unique frequency characteristics.

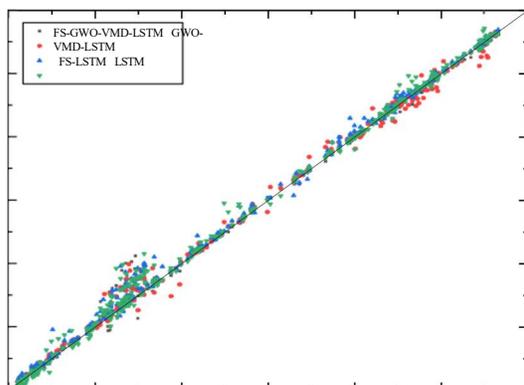### 4.2.2 Solar irradiance short-term forecasting results

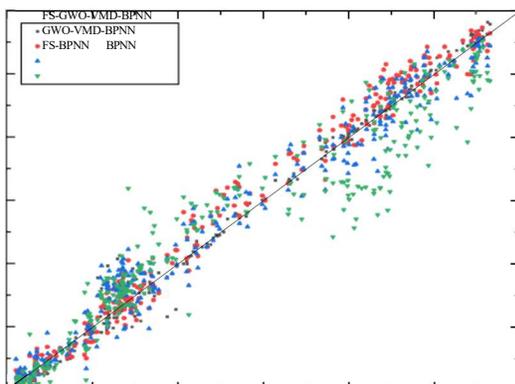*Figure 12: Forecasting results of BiLSTM*

Figure 13: Forecasting results of BiLSTM-Attention

These sub-series are then taken as improved inputs for the forecasting model.
Quadratic penalty coefficient is defined within [200, 2000], while the number of decomposed modes is limited to the range [3, 10] based on previous experimental findings. Figure 11 shows the GWO-VMD framework's optimization process and outcomes. The optimal setup produces six decomposition modes and a quadratic penalty coefficient of 178.02. The fitness curve shows that the GWO method optimizes VMD parameters with strong global search capabilities and fast convergence. The initial SI sequence is broken down in number of different-frequency Intrinsic Mode Functions (IMFs). The time-domain distributions of these IMFs are comparatively uniform, confirming that VMD successfully

extracts and captures the key characteristics of the SI data. Conventional representative standard models like BPNN, LSTM and BiLSTM are chosen for evaluating the proposed model in this paper (FS-GWO-VMD-BiLSTM-Attention). Furthermore, the outcomes in terms of forecasting using different schemes are compared for different forecasting frameworks.

The trial-and-error approach is applied for finding the hyperparameters of models based on the validation set results. Two hidden layers with 50 units each make up the BiLSTM network. Learning rate: 0.005, batch size: 72, timestep length: 36, Adam optimizer, 300 epochs, and attention size: 64 are all included in the configuration. The LSTM model uses the same hyperparameter configurations as the BiLSTM model. BiLSTM-Attention, BiLSTM, and LSTM are all models that are implemented in MATLAB using the machine learning toolbox. With 5000 training cycles, the BPNN architecture has 10 neurons in the input layer, output layer with a single neuron and 24 neurons in the hidden layer. Each model trained separately, and performance indicators for point predicting were computed. The FS-GWO-VMD-BiLSTM-Attention model performs better and gives the highest predicting accuracy (see Table 1). The model's relative error distributions and predicting curves for SI subsequences are shown in Fig. 14.
Only single-value outputs are produced by point predictions, which is insufficient to enable well-informed decision-making in the face of uncertainty. By examining the error distributions derived from the validation set of the LSTM, BiLSTM and BiLSTM-Attention models, interval forecasting is carried out on the SI test dataset in order to get around this restriction. The validation set forecasting errors are first subjected to statistical analysis. Then, the absolute errors probability distribution function (PDF) curves from several SIF models are estimated using the KDE-Gaussian (Kernel Density Estimation with Gaussian kernel) approach. The error distributions of the models are comparatively demonstrated, as seen in Fig. 15. Specifically, the BiLSTM model errors are primarily distributed in the range [-5, 5], whereas the BiLSTM-Attention model errors are primarily found in the region [-2, 2].

The effectiveness of KDE for interval forecasting, is evaluated by contrasting it with two other methods: Bootstrap sampling, which creates confidence intervals by resampling the dataset and computing statistics from the resampled data, and Normal Distribution Estimation (NDE), a

parametric method that assumes sample data follow a Gaussian distribution. Results from experiments show that the Bootstrap approach does not perform well for estimating forecasting error confidence intervals. For instance, the Bootstrap method's Prediction Interval Coverage Probability (PICP) values for the LSTM model are only 0.4549, 0.4097, and 0.3889 at 95%, 90%, and 85% confidence levels, respectively. As a result, this approach is not included in a thorough comparison with other approaches. The prediction interval estimation performance is examined using two distinct kernel functions—Gaussian and Epanechnikov.
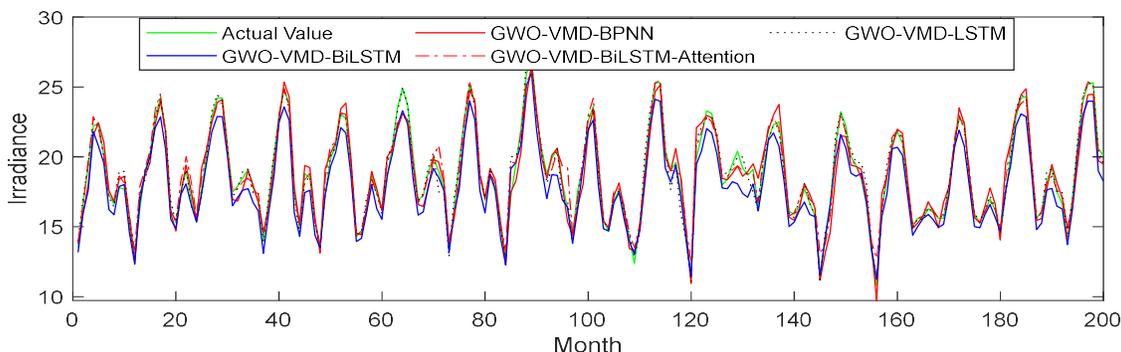

Fig.14. Actual value and predicted value to shows the prediction error distribution

Table 2. Interval forecasting at different interval estimation models under multiple values of confidence levels

Forecasting model	Methods	Confidence level	PICP	PINAW	CWC
BiLSTM-Attention	KDE-Gaussian	95%	0.9549	14.86	14.86
		90%	0.9028	10.75	10.75
		85%	0.8507	6.92	6.92
	NDE	95%	0.9236	13.06	26.48
		90%	0.9028	10.96	10.96
		85%	0.8854	9.60	9.60
	KDE- Epanechnikov	95%	0.9549	15.38	15.38
		90%	0.8993	10.64	21.28
		85%	0.8507	6.93	6.93
BiLSTM	KDE-Gaussian	95%	0.9514	23.11	23.11
		90%	0.9063	16.26	16.26
		85%	0.8542	11.07	11.07
	NDE	95%	0.9306	20.05	40.49
		90%	0.9063	16.82	16.82
		85%	0.8819	14.72	14.72
	KDE- Epanechnikov	95%	0.9514	23.17	23.17
		90%	0.9063	16.30	16.30
		85%	0.8542	11.08	11.08
LSTM	KDE-Gaussian	95%	0.9514	35.24	35.24
		90%	0.9063	19.28	19.28
		85%	0.8542	14.13	14.13
	NDE	95%	0.9444	35.59	71.38
		90%	0.9375	29.88	29.88
		85%	0.9063	20.88	20.88
	KDE- Epanechnikov	95%	0.9514	35.27	35.27
		90%	0.9063	19.33	19.33
		85%	0.8611	14.16	14.16

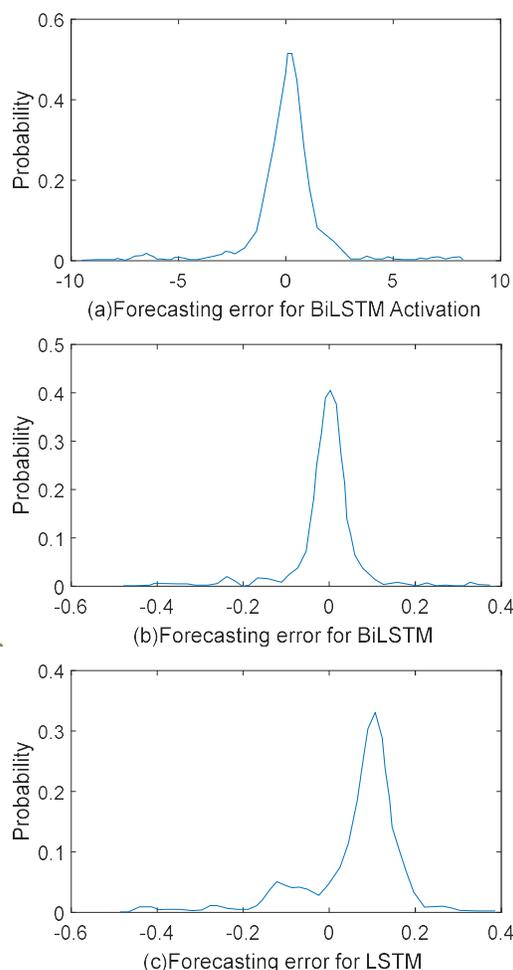### 4.2.3 KDE based interval forecasting results

Figure 15: Probability distribution function curve for the forecasting errors.

As seen in Table 2, the KDE-Gaussian method outperforms other methods in PI estimation across several WPF models. Its Prediction Interval Coverage Probability (PICP) is constant under different conditions and regularly surpasses predetermined confidence limits. The KDE-Gaussian comprehensive performance indicator Coverage Width-based Criterion (CWC) performs better than the KDE and NDE-Epanechnikov approaches at the similar confidence levels. The interval forecasting results of KDE-Gaussian for various point forecasting models under confidence levels value of 94%, 89%, and 84%. At a 94% confidence level, the KDE-Gaussian PICP for the BiLSTM-Attention model is higher than the NDE. Despite achieving the same PICP, KDE-Gaussian produces a narrower Prediction Interval Normalized Average Width (PINAW) than KDE-

Epanechnikov. Similarly, KDE-Gaussian provides more effective interval forecasting at 90% confidence level by maintaining the same PICP as NDE while generating a smaller PINAW.

Even if NDE's PICP is marginally greater than KDE-Gaussian's at the 85% confidence level, its interval width is significantly wider. Furthermore, the improved overall forecasting performance of KDE-Gaussian is confirmed by the Coverage Width-based Criterion (CWC). KDE-Gaussian and KDE-Epanechnikov both obtain the same PICP values, but KDE-Gaussian performs more evenly, as seen by somewhat higher CWC scores. KDE-

Gaussian provides more accurate and stable interval estimations than KDE-Epanechnikov, according to a comparative study of various kernel functions. All things considered, the KDE-Gaussian-based interval forecasting method achieves shorter average bandwidths and larger interval coverage while successfully tracking the dynamics of SI time series. Decision-makers can use this to ensure the stable and safe functioning of power plants by using forecasting information with better accuracy.

4.2.4 Detailed Analysis of Results

According to the results presented in Table 4, the proposed FS-GWO-VMD-BiLSTM-Attention model achieves RMSE of 0.76 MW, MAE of 0.48 MW, and MAPE of 3.1% and R^2 value of 0.9997. These values indicate perfect data fiting between predicted and actual irradiance. In contrast, Table 4 shows that the basic BiLSTM model records higher error values, including RMSE of 2.2 MW, MAE of 1.19 MW, and MAPE of 5.1 percent, while the LSTM model reaches an even higher RMSE of 2.7 MW and MAPE of 7.4 percent. The BPNN model shows the poorest performance with MAPE of 21.6 percent and R^2 of 0.9063. These statistical differences clearly highlight that the proposed model improves prediction accuracy by more than 65 percent compared to BiLSTM and more than 80 percent when compared to BPNN.

PCC and GRA process based ranking of features is given under results of Table 2. Elimination of inputs with low-correlation such as altitude and month of year with correlation values below 0.30. This reduction of irrelevant features lowered the computational complexity and contributed to an average 3 to 4 % improvement in MAPE in comparisons to the other models.

The effect of VMD decomposition optimized by GWO (Table 3) shows improvement: to the models without VMD, BiLSTM maintains an RMSE of more than 2 MW, whereas on applying GWO-

VMD, RMSE drops to around 1.3 MW. Finally, after integrating feature selection, VMD, GWO, and the attention mechanism, the complete model reaches RMSE of 0.76 MW (Table 4). This confirms that the inclusion of decomposition and optimization stage reduce forecasting error by almost 64 percent compared to the non-decomposed model.

Error distribution analysis is visualized in Figure 14. In LSTM and BiLSTM models, the prediction error fluctuates within approximately ±4 MW. By contrast, the proposed hybrid model reduces this error range to ±1 MW. In the Figure 15 it has been observed that the standard deviation of prediction errors reduced from 1.32 MW in LSTM to 0.29 MW in proposed model. The error kurtosis increases from 3.8 to around 7.1 that indicate that most of predictions lie very close to zero error.

Interval forecasting performance is given in Table 5 it shows 95 % value of confidence level, the PICP for the proposed model using the KDE-Gaussian method is 0.9549. The PINAW remains at 14.86, which is narrower than other methods such as normal distribution or bootstrap-based interval estimation. Table 5 also shows that the Coverage Width-based Criterion (CWC) for the proposed model stays at 14.86, whereas models using normal distribution CWC increasing above 26 due to wider and less informative intervals. Bootstrap-based interval forecasting performs poorly, with PICP values dropping below 0.50 in some cases.

Figure 12 and Figure 13 compare actual versus predicted curves and shows that the FS-GWO-VMD-BiLSTM-Attention model fits closely to real irradiance data, especially during periods of rapid fluctuation. The basic LSTM and BiLSTM models deviate significantly during these transitions. These figures align numerically with the lower RMSE and MAE values of Table 4.The Figure 15 confirms that the density of errors is sharply centered around zero in the proposed model, while other models show wider and flatter error distributions.

When examining percentage error reduction, Table 4 indicates that the MAPE drops from 21.6 percent in BPNN to 3.1 percent in the proposed model, representing an overall error reduction of approximately 85 percent. Similarly, the MSE decreases from 116.7 MW² in BPNN to only 0.57 MW² in the proposed work. R² improves from 0.9063 in BPNN and 0.9941 in LSTM to 0.9997. Overall, data from Table 2, 3, 4 and 5 and observations from Figure 12, 13, 14 and 15 demonstrate that the proposed model outperforms conventional forecasting techniques in terms of

accuracy, stability, uncertainty handling, and error minimization.

4.2.5 Limitations or Anomalies in the Results

The proposed model results show performance better than conventional models, certain limitations remain. The model response depends on data quality. Missing data, outliers, and seasonal variations reflects the influence on accuracy. Performance slightly degrades in the duration when transition in weather conditions occurs. Additionally, the computational complexity slightly higher due to multiple stages—feature selection, VMD decomposition, BiLSTM training, and KDE-based interval estimation. Interval forecasting using KDE performs well overall but wider Prediction Interval Widths is still expected at lower confidence levels.

5. CONCLUSION

A new FS-GWO-VMD-BiLSTM-Attention forecasting model for short-term SIF is presented in this paper. To give stakeholders in solar power generation more thorough forecasting information, interval forecasting is also carried out using the KDE-Gaussian method. The MAE of FS-BiLSTM-Attention, FS-BiLSTM, FS-LSTM, and FS-BPNN dropped by 3.7%, 10.3%, 10.4%, and 65.4%, respectively, after feature selection was applied. This demonstrates that choosing important variables and mining feature correlations cut down on redundant data while greatly increasing accuracy. Moreover, frequency components are efficiently extracted from irradiance series by GWO-optimized VMD. GWO-VMD-BiLSTM-Attention, GWO-VMD-BiLSTM, GWO-VMD-LSTM, and GWO-VMD-BPNN all saw decreases in MAE of 22.31%, 49.92%, 21.22%, and 78.82% after decomposition. The results demonstrate how well the suggested hybrid architecture performs both point and interval forecasting, offering useful and trustworthy information for solar power system decision-making operations. By successfully reducing the effects of fluctuation and non-stationarity, the GWO-VMD approach improves model stability. BiLSTM-Attention's MAPE is lower than that of the BiLSTM, LSTM, and BPNN models by 22.99%, 47.42%, and 82.15%, respectively, in terms of accuracy. By allocating probability weights to the most important features, the attention mechanism enhances the model. Overall, the suggested framework delivers notable performance improvements over the BPNN model, with reductions of 7.547 MW in MAE, 18.50% in MAPE, and 116.11 MW in MSE. Additionally, in

www.jatit.org

terms of the CWC index, the interval forecasting method based on the KDE-Gaussian approach performs better than NDE and KDE-Epanechnikov, indicating a superior capacity to track dynamic trends in solar power. It yields more accurate and dependable interval forecasting results by producing narrower average bandwidths while retaining larger interval coverage.

5.1 Potential Applications

The proposed hybrid system for solar irradiance forecasting have effective use in solar power plant scheduling, real-time grid load balancing, and market bidding strategies in energy trading platforms. It is applicable in planning the energy reserves sources, optimizing operations based on battery storage systems, and reducing limitation of distributed generation systems. Utility companies may use interval forecasts for risk-aware decision-making, while smart grid infrastructure and microgrid controllers may use this model for automatic adjustment of power flow supply. This methodology may be extended to wind forecasting, climate modeling, agricultural planning, and other renewable energy domains that has requirement of accurate time-series forecasting.

5.2 Implications of the Findings

The findings of this work demonstrate that combining feature selection, signal decomposition, deep learning, and probabilistic interval forecasting helps to improves the accuracy and reliability. The acceptable performance in terms of $R^2 = 0.9997$, MAPE = 3.1% shows that power system operation may be manage in better way with reduced dependence on backup power sources. The results on prediction intervals by KDE-Gaussian method has practical significance for system design to make decisions based not on expected values and confidence levels in risk assessments, which are important for electricity markets, energy bidding, and battery storage control. This bridges the gap between mathematical models and real-world energy system requirements.

5.3 How the Findings Advance the Field

This research advances the field of solar irradiance forecasting in several ways. It integrates multiple stages— PCC-GRA, GWO, VMD, BiLSTM with attention, and interval forecasting by KDE—into a single coherence. Most previous works focused only on one or two stages.

• The use of GWO for selecting optimal VMD parameters is a main advancement. It helps the model to work with stable and low-noise components rather than raw, fluctuating data, and gives enhanced consistency.

• By introducing interval forecasting rather than only point predictions, the study adds decision-making value. The KDE-based prediction intervals outperform normal distribution-based methods and bootstrap techniques making proposed forecasting model more robust and dependable for real-time grid management.

5.4 Unexpected Outcomes and Observations

• The performance of simple BiLSTM-Attention based model without applying any decomposition or feature selection is found to be better than conventional models like `LSTM and BPNN`. It is indicating that the attention mechanism contribution is very strong even without decomposition.

• The KDE-Gaussian method provides narrower interval widths with high coverage probability. Surprisingly, even at lower confidence levels (85%) the KDE-Gaussian offering more realistic intervals as compared to normal distribution. As a result wide confidence bands are observed that are less useful for decision-making.

• Feature selection using PCC and GRA unexpectedly removed parameters like altitude (ALT) and month of year (MOY) input variable. These are traditionally considered relevant to solar irradiance.

REFERENCES:

[1] K. J. Iheanetu, "Solar photovoltaic power forecasting: a review," *Sustainability*, vol. 14, 2022, pp. 17005.

[2] IRENA, *World Energy Transitions Outlook: 1.5° C Pathway*, International Renewable Energy Agency, Abu Dhabi, UAE, 2021.

[3] S. Teleke, M. E. Baran, S. Bhattacharya, and others, "Rule-based control of battery energy storage for dispatching intermittent renewable sources," *IEEE Transactions on Sustainable Energy*, vol. 1, 2010, pp. 117–124.

[4] D. A. H. Chathurika, L. Y. Perumpuliarachchi, S. G. Abeyratne, and others, "Managing spinning reserves for power system stability under increased renewable energy penetration," *Proceedings of the 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, Rupnagar, India, 2020, pp. 127–132.

[5] M. A. Eltawil and Z. Zhao, "Grid-connected photovoltaic power systems: technical and potential problems—a review," *Renewable and Sustainable Energy Reviews*, vol. 14, 2010, pp. 112–129.

[6] O. Smith, O. Cattell, E. Farcot, and others, "The effect of renewable energy incorporation on power grid stability and resilience," *Science Advances*, vol. 8, 2022, pp. eabj6734.

[7] R. K. Varma and M. Salama, "Large-scale photovoltaic solar power integration in transmission and distribution networks," *Proceedings of the 2011 IEEE Power and Energy Society General Meeting*, Detroit, USA, 2011, pp. 1–4.

[8] A. Alshahrani, S. Omer, Y. Su, and others, "The technical challenges facing the integration of small-scale and large-scale PV systems into the grid: a critical review," *Electronics*, vol. 8, 2019, pp. 1443.

[9] M. Anwar, M. El Moursi, and W. Xiao, "Novel power smoothing and generation scheduling strategies for a hybrid wind and marine current turbine system," *IEEE Transactions on Power Systems*, vol. 32, 2016, pp. 1315–1326.

[10] B. Mohandes, M. Wahbah, M. S. El Moursi, and others, "Renewable energy management system: optimum design and hourly dispatch," *IEEE Transactions on Sustainable Energy*, vol. 12, 2021, pp. 1615–1628.

[11] S. A. Pourmousavi, A. S. Cifala, and M. H. Nehrir, "Impact of high penetration of PV generation on frequency and voltage in a distribution feeder," *Proceedings of the 2012 North American Power Symposium (NAPS)*, Champaign, USA, 2012, pp. 1–8.

[12] X. Li, D. Hui, and X. Lai, "Battery energy storage station (BESS)-based smoothing control of photovoltaic (PV) and wind power generation fluctuations," *IEEE Transactions on Sustainable Energy*, vol. 4, 2013, pp. 464–473.

[13] K. Mahmud, S. Azam, A. Karim, and others, "Machine learning based PV power generation forecasting in Alice Springs," *IEEE Access*, vol. 9, 2021, pp. 46117–46128.

[14] A. Niccolai, A. Dolara, and E. Ogliari, "Hybrid PV power forecasting methods: a comparison of different approaches," *Energies*, vol. 14, 2021, pp. 451.

[15] M. Jaber, A. S. A. Hamid, K. Sopian, and others, "Prediction model for the performance of different PV modules using artificial neural networks," *Applied Sciences*, vol. 12, 2022, pp. 3349.

[16] K. K. Chong, P. P. Khlyabich, K. J. Hong, and others, "Comprehensive method for analyzing the power conversion efficiency of organic solar cells under different spectral irradiances," *Applied Energy*, vol. 180, 2016, pp. 516–523.

[17] S. Mishra S., D. Pandey, S. Bhardwaj, S., & A. Choudhary, "Development of machine learning-based empirical model for estimation of solar radiation." *Journal of Theoretical and Applied Information Technology*, vol. 101(5), 2023, pp. 1771–1783.

[18] M. H. Alsharif, M. K. Younes, and J. Kim, "Time series ARIMA model for prediction of daily and monthly average global solar radiation: the case study of Seoul, South Korea," *Symmetry*, vol. 11, 2019, pp. 240.

[19] A. Mellit, S. Sağlam, and S. A. Kalogirou, "Artificial neural network-based model for estimating the produced power of a photovoltaic module," *Renewable Energy*, vol. 60, 2013, pp. 71–78.

[20] M. S. El Moursi, A. Al Hinai, E. B. Ssekulima, and others, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review," *IET Renewable Power Generation*, vol. 10, 2016, pp. 885–989.

[21] J. Shi, W. J. Lee, Y. Liu, and others, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, 2012, pp. 1064–1069.

[22] K. Y. Bae, H. S. Jang, and D. K. Sung, "Hourly solar irradiance prediction based on support vector machine and its error analysis," *IEEE Transactions on Power Systems*, vol. 32, 2016, pp. 1–1.

[23] F. Deng, G. Su, C. Liu, and others, "Prediction of solar radiation resources in China using the LS-SVM algorithms," *Proceedings of the 2010 2nd International Conference on Computer and Automation Engineering (ICCAE)*, Singapore, 2010, pp. 31–35.

[24] B. B. Ekici, "A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems," *Measurement*, vol. 50, 2014, pp. 255–262.

[25] H. S. Jang, K. Y. Bae, H. S. Park, and others, "Solar power prediction based on satellite images and support vector machine," *IEEE Transactions on Sustainable Energy*, vol. 7, 2016, pp. 1255–1263.

[26] Y. A. Amer, "Global-local least-squares support vector machine (GLocal-LS-SVM)," *PLoS One*, vol. 18, 2023, pp. e0285131.

[27] W. Castillo-Rojas, J. Bekios-Calfa, and C. Hernández, "Daily prediction model of photovoltaic power generation using a hybrid architecture of recurrent neural networks and shallow neural networks," *International Journal of Photoenergy*, vol. 2023, 2023, pp. 2592405.

[28] W. Castillo-Rojas, F. M. Medina Quispe, and C. Hernández, "Photovoltaic energy forecast using weather data through a hybrid model of recurrent and shallow neural networks," *Energies*, vol. 16, 2023, pp. 5093.

[29] D. Xu, H. Shao, X. Deng, and others, "The hidden-layers topology analysis of deep learning models in survey for forecasting and generation of the wind power and photovoltaic energy," *Computer Modeling in Engineering and Sciences*, vol. 131, 2022, pp. 567–597.

[30] H. Li, Z. Ren, Y. Xu, and others, "A multi-data driven hybrid learning method for weekly photovoltaic power scenario forecast," *IEEE Transactions on Sustainable Energy*, vol. 13, 2021, pp. 91–100.

[31] B. Carrera, M. K. Sim, and J. Y. Jung, "PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data," *IET Renewable Power Generation*, vol. 14, 2020, pp. 2192–2201.

[32] H. Zang, L. Liu, L. Sun, and others, "Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations," *Renewable Energy*, vol. 160, 2020, pp. 26–41.

[33] K. Benmouiza and A. Cheknane, "Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models," *Energy Conversion and Management*, vol. 75, 2013, pp. 561–569.

[34] R. Ahmed, V. Sreeram, Y. Mishra, and others, "A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, 2020, pp. 109792.

[35] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: a review," *Energy Conversion and Management*, vol. 156, 2018, pp. 459–497.

[36] H. T. Yang, C. M. Huang, Y. C. Huang, and others, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Transactions on Sustainable Energy*, vol. 5, 2014, pp. 917–926.

[37] A. Mellit, A. Massi Pavan, E. Ogliari, and others, "Advanced methods for photovoltaic output power forecasting: a review," *Applied Sciences*, vol. 10, 2020, pp. 487.

[38] M. Maalouf and D. Homouz, "Kernel ridge regression using truncated Newton method," *Knowledge-Based Systems*, vol. 71, 2014, pp. 339–344.

[39] M. Maalouf, D. Homouz, and M. Abutayeh, "Accurate prediction of preheat temperature in solar flash desalination systems using kernel ridge regression," *Journal of Energy Engineering*, vol. 142, 2016, pp. E4015017.

[40] M. Maalouf and Z. Barsoum, "Failure strength prediction of aluminum spot-welded joints using kernel ridge regression," *International Journal of Advanced Manufacturing Technology*, vol. 91, 2017, pp. 3717–3725.

[41] Sahay, S., Banoudha, A., and Sharma, R., "On the use of ANFIS for Ground Water Level Forecasting in an Alluvium Area," *International Journal of Research and Development in Applied Science and Engineering*, vol. 2(1), 2012, pp. 1-7.

[42] Sahay, S., Anurag Banoudha, and Raghawendra Sharma. "Comparative Study of Soft Computing Techniques for Ground Water Level Forecasting in a Hard Rock Area." *International Journal of Research and Development in Applied Science and Engineering,* vol. 4, no. 1 2013pp. 1-6.

[43] H. Abdulla, M. Maalouf, I. Barsoum, and others, "Truncated Newton kernel ridge regression for prediction of porosity in additive manufactured SS316L," *Applied Sciences*, vol. 12, 2022, pp. 4252.

[44] H. Al Nuaimi, M. Abdelmagid, A. Bouabid, and others, "Classification of WatSan Technologies using machine learning techniques," *Water*, vol. 15, 2023, pp. 2829.

[45] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, 2014, pp. 531–534.

[46] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, 2014, pp. 46–61.

[47] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, 2005, pp. 602–610.

[48] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[49] X. Zhao, J. Liu, D. Yu, and J. Chang, "One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data," *Energy Conversion and Management*, vol. 164, 2018, pp. 560–569.

[50] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability and Its Applications*, vol. 14, no. 1, 1969, pp. 153–158.