Energy Conversion and Management: X 20 (2023) 100486

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

Resource efficient PV power forecasting: Transductive transfer learning based hybrid deep learning model for smart grid in Industry 5.0

Umer Amir Khan [a], Noman Mujeeb Khan [b], Muhammad Hamza Zafar [c,*]

[a] Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[b] Department of Electrical and Electronics Engineering, Beaconhouse International College, Islamabad 46000, Pakistan
[c] Department of Engineering Sciences, University of Agder, Grimstad 4879, Norway

ARTICLE INFO

Keywords:
Deep learning models
Renewable energy
Industry 5.0
Photovoltaic energy systems

ABSTRACT

This paper presents an innovative approach for enhancing power output forecasting of Photovoltaic (PV) power plants in dynamic environmental conditions using a Hybrid Deep Learning Model (DLM). The hybrid DLM employs a synergy of Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) network, and Bidirectional LSTM (Bi-LSTM), effectively capturing spatial and temporal dependencies within weather data crucial for accurate predictions. To optimize the DLM's performance efficiently, a unique Kepler Optimization Algorithm (KOA) is introduced for hyperparameter tuning, drawing inspiration from Kepler's laws of planetary motion. By leveraging KOA, the DLM attains optimal hyperparameter configurations, elevating power output prediction precision. Additionally, this study integrates Transductive Transfer Learning (TTL) with the deep learning models to enhance resource efficiency. By leveraging knowledge gained from previously learned tasks, TTL enables the DLM to improve its forecasting capabilities while minimizing resource utilization. Datasets encompassing environmental parameters and PV plant-generated power across diverse sites are employed for DLM training and testing. Three hybrid models, amalgamating KOA, CNN, LSTM, and Bi-LSTM techniques, are introduced and evaluated. Comparative assessment of these models across distinct PV sites yields insightful observations. Performance evaluation, focused on short-term PV power forecasting, underscores the hybrid DLM's superiority over individual CNN and LSTM models. This hybrid approach achieves remarkable accuracy and resilience in predicting power output under varying weather conditions, showcasing its potential for efficient PV power plant management.

1. Introduction

The modern world is laid on the foundation of electrical energy. All the human activities revolve around technologies being run by electricity. From the comfort of our home to national security, from food security to health, from economic growth to human development, all these aspects are directly linked with the availability of Electrical Energy. Conventionally, humans have exploited natural resources to generate electricity for decades. However, the exponential demand of electricity has resulted in dilapidation of non-renewable resources and also severely damaged the environment [1]. Global warming is a serious threat converting into reality which is threatening life of millions of humans in coming decades. Therefore, it is very important for researchers to provide solutions which can reverse the process of environmental degradation along with sustainably supporting the growth of technology driven global development.

One of the promising solutions which can contribute in decreasing the dependency on fossil fuel based CES is generating electricity from renewable energy sources like Photovoltaics (PV) and Wind Energy. Last two decades have seen an enormous integration of RES with conventional electricity networks throughout the world [2]. Due to large scale integration of RES in power grids, the power utility companies are facing the problem of optimal power dispatch due to the intermittent behavior of RES [3].

The variations in RES power being injected into a power grid are directly linked with the changing environmental conditions which are beyond human control. On the contrary, CES power can be exactly controlled and predicted through out its operation. This property of CES enables utilities to economically dispatch power into the grid. Economic dispatch involves considering the variable costs of each CES power plant along with the anticipated energy market prices. A CES power plant will

* Corresponding author.
E-mail address: muhammad.h.zafar@uia.no (M.H. Zafar).

https://doi.org/10.1016/j.ecmx.2023.100486

Available online 22 November 2023

Nomenclature

RES	Renewable Energy Sources
CNN	Convolutional Neural Network
Bi-LSTM	Bi-Directional Long Short Term Memory
KOA	Kepler Optimisation Algorithm
NMSE	Normalized Mean Square Error
MAE	Mean Absolute Error
RE	Relative Error
TKCL	Transductive Transfer Learning assisted Kepler Optimization Algorithm-based Convolutional Neural Network with Long Short Term Memory Network
SCADA	Supervisory Control and Data Acquisition
CES	Conventional Energy Sources
LSTM	Long Short Term Memory
DLM	Deep Learning Model
KCBL	Kepler Optimization Algorithm-based Convolutional Neural Network with Bi-Long Short Term Memory Network
RMSE	Root Mean Square Error
R^2	R Square
TKCBL	Transductive Transfer Learning based KCBL
TKL	Transductive Transfer Learning assisted Kepler Optimization Algorithm-based Long Short Term Memory Network
GCT	Granger Causality Test

only operate if it's cost of power production is lower than the revenue generated through sales of that power. Economic dispatch involves creating a schedule that determines the allocation of available capacity for the CES power plants which is based on predetermined knowledge of how much power output can be attained from a certain CES power plant at a certain time. Thus the changing load demand can be managed efficiently through fixed inputs exactly meeting the load demand [4]. Unfortunately, the intermittent behavior of RES makes this scheduling very complex as the input to power grids are now dynamic depending on environmental conditions and not directly dependant on load demand [5]. Therefore, Smart Grids having both CES and RES require accurate short term and long term power forecasting to efficiently balance the dynamic input to dynamic load for sustainable operation. Power Grids of the future will be power systems having dynamic input and they have to produce dynamic outputs. Power forecasting and Load forecasting are the techniques that are generally used in the literature [6]. Load forecasting deals with predicting expected load demand in both the near and far future. Power forecasting is the technique used to predict how much power can be generated in the future. To meet the load demand, we must have clear identification of power forecast for the future so that we are able to balance the demand and supply of power at the time of power dispatch. In case of a deficiency of local RES, we might have to supplement the load demand by CES, or purchase RES based electricity from other producers in the region [7]. Fig. 1 shows a layout of a dynamic power system with varying input and output conditions. Due to dependence on environmental conditions making forecasting complex and tedious, intelligent power dispatch techniques are required to maintain the equilibrium between dynamic supply and demand.

The nonlinear behavior in the environmental conditions makes the power forecasting a real challenge [8]. For photovoltaics, clouds, rain storms, wind gusts, partial and full shading, snow fall etc. are the few conditions that can cause drastic fall in the output of the PV plants. Similarly, for Wind Turbines, the variability in the wind speeds can be a real challenge to correctly forecast the power output that will be available from the RES. The power forecasting can be further divided into long-duration, medium-duration and short-duration [9]. Long-duration forecasting covers a time period between a single month up to a year ahead forecasting. This allows long-term decision making for the planning and resource management in any power system containing a considerable amount of RES penetration. The second type of

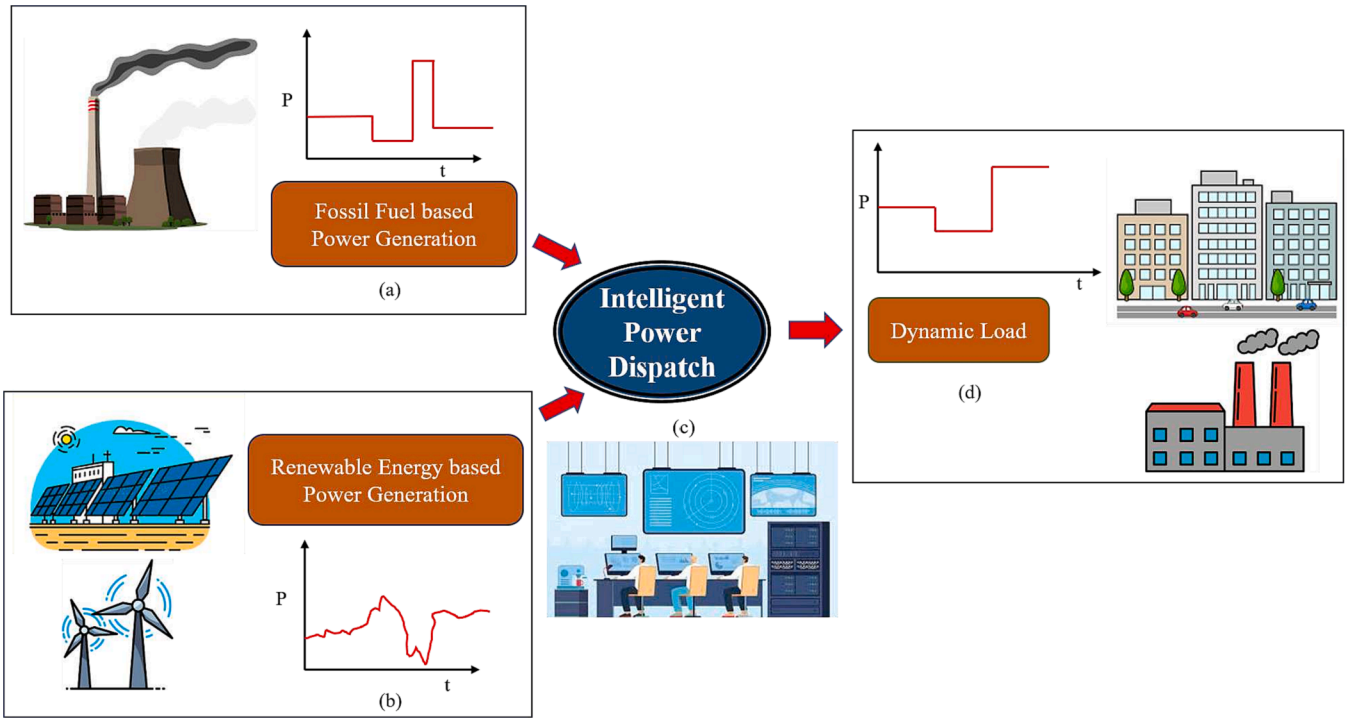

Fig. 1. Dynamic power sytems with changing input and output power (a) Conventional Fossil fuel based power plants which can produce constant power as per requirement (b) Intermittent output from the RES (c) Intelligent power dispatch control system with Machine Learning based power forecasting (d) Dynamic load as per consumer requirement.

forecasting is middle-duration forecasting which covers a week ahead forecasting up to one month which effectively assist in economic dispatch support for power system planning and optimization. Short-duration forecasting covers a week or less than a week power forecasting and plays a critical role in meeting the dynamic load demand in a power system. Day ahead forecasting and hourly forecasting are sub-categories of short-duration forecasting which plays a critical role in economic dispatch, real-time adjustments, and catering for dynamic loads in power systems having significant RES penetration [10].

It is the short-duration forecasting which plays a fundamental role in electricity markets and is being used for balancing the dynamic supply and dynamic demand in real time. In this research work we are focusing on short-duration PV power forecasting starting from day ahead to final hour of operation. The power forecasting models are generally divided into physical [11] and statistical-methods [12]. Physical methods utilize mathematical models to forecast the weather conditions. This technique performs well for stable weather conditions but the accuracy of power forecasting is greatly reduced when considering variable weather scenarios [13]. The second technique uses statistical methods based on machine learning techniques. These techniques use the relationship between weather patterns with the power output of RES plant developed. These correlations are formed during the learning process of forecasting models which are developed using historical data from the RES power plants. Datasets of historical data play a key role in the accuracy of forecasting through machine learning models. Few commonly used statistical methods include gray wolf optimization technique [14], Markov Chain [15], Fuzzy Logic [16], Regression and Auto-Regression techniques [17].These methods show satisfactory performance if they are optimized continuously and are trained on relatively large and good quality datasets. However, with the introduction of non-linearity and complex datasets, the accuracy of these machine learning techniques are severely affected.

1.1. Contributions and Paper Organization

In this paper, Hybrid DLM is proposed for output power estimation of PV sites under dynamic environmental conditions. The proposed structure of the Hybrid DLM based PV Power forecasting is elaborated in Fig. 2. First, the data is collected from different PV sites which includes the environmental parameters and the generated power of the PV plant. After that, the data pre-processing is done which gives us useful input features i.e. global irradiance, total irradiance, air temperature, air pressure, and relative humidity. After processing and normalization, the dataset is divided into training and testing parts. The proposed hybrid DLM is trained on the training dataset and to check the effectiveness of the model, the trained model is then tested on testing data. The short-term PV power forecasting performance is then verified by carrying out statistical analysis.

The contributions of this paper are listed below:

- **Comprehensive Analysis of Hybrid Models:** A detailed examination of three distinct hybrid deep learning models designed for PV power forecasting is presented. These models combine CNN, LSTM and Bi-LSTM architectures to effectively capture spatial and temporal dependencies in weather data.
- **Innovative Kepler Optimization Algorithm (KOA) Implementation:** A novel approach is introduced, known as Kepler Optimization Algorithm-based Convolutional Neural Network with Bi-Long Short Term Memory Network (KCBL). This architecture harnesses the power of KOA for efficient hyperparameter tuning, resulting in improved forecasting accuracy.
- **Efficient Hyperparameter Tuning using KOA:** The paper demonstrates the successful utilization of the Kepler Optimization Algorithm for precise tuning of hyperparameters in the CNN-Bi-LSTM model. This approach significantly enhances the efficiency of PV power forecasting.
- **Comparative Analysis of Models:** A comprehensive comparison is conducted between various models, including KOA Algorithm-based CNN-LSTM (KOA-C-LSTM) and KOA Algorithm-based LSTM Model (KOA-LSTM). This analysis provides insights into the strengths and weaknesses of each model in the context of PV power forecasting.
- **Integration of Transductive Transfer Learning:** The paper introduces the application of Transductive Transfer Learning, a method that leverages knowledge gained from prior tasks to enhance

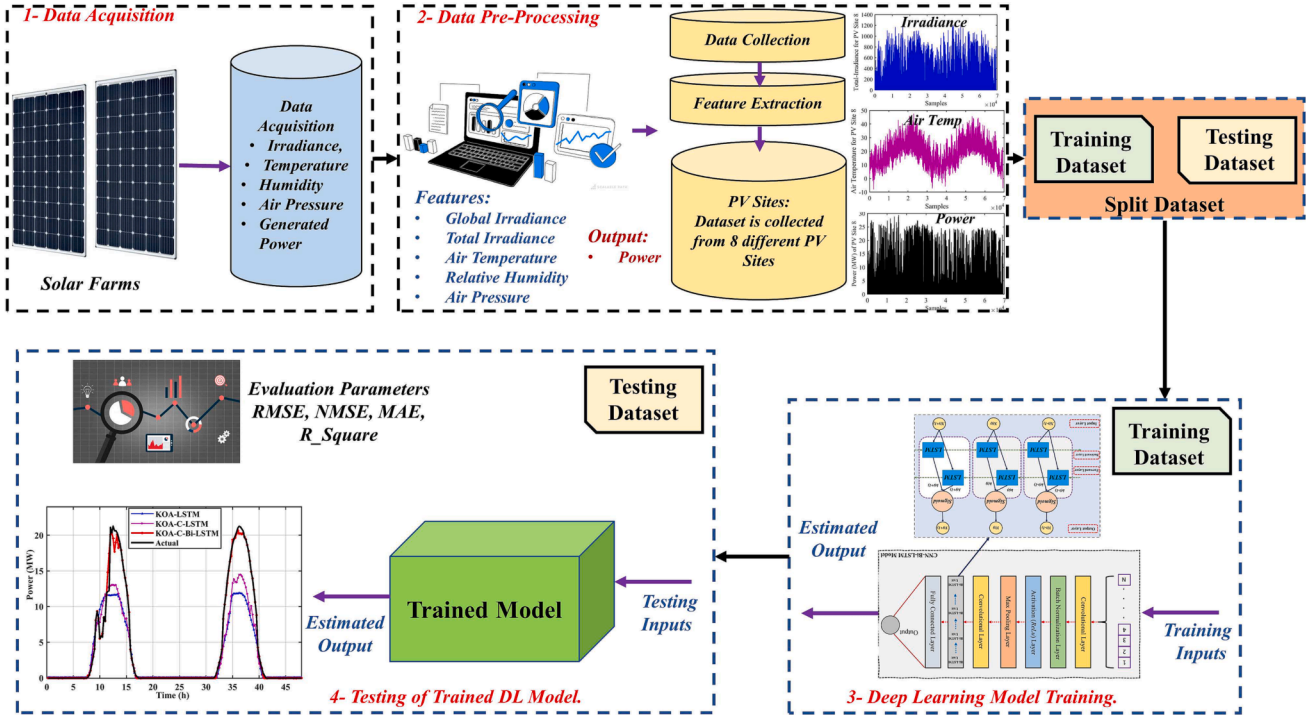

Fig. 2. Proposed Flow for Deep Learning-based PV Power Forecasting Model.

the forecasting capabilities of the models, while maintaining resource efficiency.

- **Multi-Site Dataset and Testing:** The proposed hybrid model is trained and rigorously tested on datasets from eight different PV sites, collected under diverse environmental conditions. This testing ensures the robustness and versatility of the model's forecasting performance.
- **Evaluation Metrics for Comparison:** To facilitate comparison, the paper presents evaluation metrics such as NMSE, RMSE, MAE, R^2, and RE. These metrics provide a comprehensive assessment of the models' performance.
- **Efficient Power Forecasting:** Through empirical analysis, it is demonstrated that the proposed TKCBL model consistently excels in efficiently forecasting PV power across datasets from different PV sites.

2. Industry 5.0 and Power Forecasting Applications

The modern world and global economy finds its foundation in Industrial Revolutions that have occurred in last two centuries [18]. The first industrial revolution started around 1780 with the advent of steam engine and its utilization in industrial activities; the second industrial revolution is marked with the inclusion of assembly lines and mass output in factories; and the third industrial revolution utilized the concept of automation and electronic based control which led to the development of controlled industrial processes and development of software industry and IT to assist in production. The Fourth Industrial Revolution (Industry 4.0) is marked with integration of Internet of Things (IoT) and the cloud computing to develop cyber-physical networks with real-time connectivity, and control over various industrial and productive activities. Unfortunately, from the beginning of First Industrial Revolution, conservation of environment has never been the priority of global industrialists [19].

Due to rapid and large scale industrialization, our planet earth has suffered immensely due to environmental emissions. The greenhouse gas emissions due to electrical power systems hold a significant share in the global warming which is severely effecting the planet. Sadly, relatively recently launched Industry 4.0, despite backed with latest developments in technology, has not been centered around environmental protection and dealing with the harmful affects of global warming. Keeping this thing in mind, within a relatively shorter period after the launch of Industry 4.0, the concept of Industry 5.0 is launched to cope up with the challenges being faced by the whole humanity due to global warming and environmental deterioration. The Industry 5.0 is primarily a human-centric strategy with three fundamental pillars: people-oriented, environment-oriented, and resilience oriented [20,21]. Backed up with the power of Artificial Intelligence and Machine learning, it is expected that the existing electrical power systems which are the major contributor in greenhouse gas emissions and global warming will be converted into more green and sustainable energy systems. Machine learning based Power Forecasting can play a significant role in achieving this target.

Electrical power grid which is clean, sustainable, and smart is expected to be the most critical human advancement of 21st century which can ensure a technology driven future. This intelligent electrical power grid is termed as Smart Grid [22,23]. One of the fundamental principles which is at the core of Smart Grid is about being consumer centered instead of producer centered. This principle matches with first basic pillar of Industry 5.0 that is people-centered. Conventional power grids are based on the principle of passive consumption from the consumer with no say in how the power is being generated. The prosumers (producers and consumers) of the Smart Grid will be more involved in the power being generated and consumed. The electricity users will be billed according to the dirty or clean energy consumption through the concept of real time and dynamic pricing [24]. [25] The utility can implement this concept when it can accurately forecast the power and confidently convey it to the users beforehand. To enable the consumers to have this flexibility, efficient power forecasting algorithms and techniques using Artificial Intelligence and Machine learning are imperative to be integrated into Smart Grid. The proposed research work addresses the issue of correct power forecasting and is thus linked with Smart Grid and implementation of Industry 5.0.

Industry 5.0 standards include the concept of resilience. Resilience in Smart Grid aims at ensuring uninterruptible and high quality power despite any kind of environmental, physical, and cyber threats which can disrupt the power systems [26,27]. With increasing penetration of RES in power grids all over the world, the power flow is getting more and more dependent on short term to long term environmental conditions. The variations in environmental condition will change the output of RES and can severely affect the balance of supply and demand in the power system. Correct and efficient power forecasting will allow power grid operators to foresee and tackle the power imbalances and provide resiliency to the grid under these conditions. The power forecasting technique presented in this research work aims at directly contributing in increasing the resiliency of the power systems.

Industry 5.0 also targets to bring the attention back towards our Environment and repercussions of global warming. Industry 5.0 aims at improving the global environment and decrease the negative impacts of the existing power grid by enabling maximum integration of RES with the power grid and less and less dependence on fossil-fuel based power generation units. The only way a considerable amount of conventional power plants can be replaced with RES is through intelligent, accurate and quick power forecasting and predicting the power output for dynamic dispatch. The proposed research work presents a novel and unique Machine Learning based power forecasting technique to predict power output of large scale RES integration in the grid. Fig. 3 summarizes the basic concept of how Power Forecasting in Smart Grids will be implementing the concepts of Industry 5.0 in electricity power systems.

3. PV Power Datasets from Field Sites

3.1. Dataset Acquisiton

The availability of diverse and real-world data from the field stations is imperative for the correct training of machine learning algorithms which can then by effectively used in real-world applications. Solar stations often employ a SCADA system, which integrates hardware and software components to monitor, control, and analyze various processes [28]. These processes include data visualization, alarm functions, fault detection, and emergency offloading which are also remotely monitored [29]. This study utilizes datasets collected from eight solar stations using SCADA systems and are in public domain [30]. Table 1 provides the basic information about the solar sites including the nominal output capacity, the PV Panels used and total number of panels available. The selected solar stations had capacities ranging from 30 MW to 130 MW. The data were accessed through a remote monitoring platform and downloaded as.xlsx files, with authorized access granted by the respective owners.

The selected solar station sites were located in North, Central, and Northwest China to ensure the representation of diverse climate zones and geographic locations. These sites encompassed various terrains, including deserts, mountains, and plains. It is important to note that all original datasets were obtained from the Chinese State Grid, a third-party organization, and the data collection process was beyond the control of the authors. The variation of different features and output power is shown in Fig. 4.

3.2. Data Pre-Processing

The available datasets are not ideal and contains missing data as expected from real world datasets. These irregularities can be due to various factors such as sensor malfunction and communication issues.

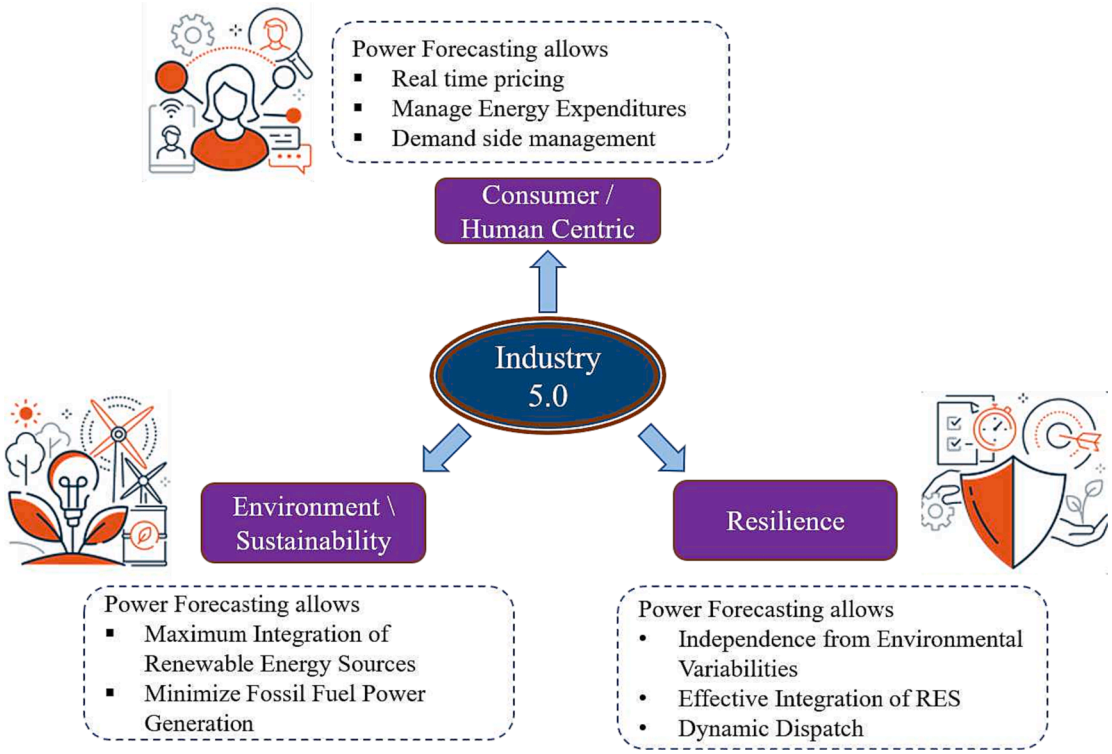

Fig. 3. The three fundamental attributes of Industry 5.0 and its relationship with power forecasting application.

Table 1
Technical Specs of the PV systems used in all PV sites.

Solar Site Name	Nominal Cap. (MW)	PV Panel Model	No. of Panels
PV Site 1	50	NA	NA
PV Site 2	130	NA	NA
PV Site 3	30	CS6U-325P	27995
PV Site 4	130	NA	NA
PV Site 5	110	JNMP60-255	36828
PV Site 6	35	SUN2000-50KTL-C	703
PV Site 7	30	NA	60
PV Site 8	30	CS6K-260P-PG/CS6K-255P-PG	5986/25383

The missing data was indicated by null, NA, 0.001, and −0.99 within the datasets. To ensure the accuracy and reliability of our proposed model, handling missing data and minimizing the impact of noise and outliers is essential.

Two different data pre-processing techniques were employed to address these issues, the linear interpolation imputation method and the K-nearest neighbors (KNN) imputation method [31]. The first approach involves estimating the missing values by calculating the average of the neighboring data points. We preprocessed the raw PV sites data sets mentioned in Table 1. Similarly, we applied the K-nearest neighbors (KNN) imputation technique to process the same raw three PV sites data sets mentioned above. KNN imputation fills in missing values based on the values of the K most similar neighboring data points. In this study, K with value of 5 is used to improve the accuracy of preprocessing. The performance of both approaches is discussed in the result section by utilizing both processed sets of data to train and test our proposed forecasting models.

3.3. Dataset Normalization

After preprocessing, Normalization is a crucial step in preparing data for DLM, including PV power forecasting DLMs. It ensures that input features are standardized and have a consistent range, which aids in model convergence, stability, and overall performance. One popular normalization technique is min–max normalization, which scales the data to a specific range, typically between 0 and 1 [32].

Min–max normalization transforms each data point x_i in a feature vector X to a normalized value x_i' using the following equations:

$$x_i' = \frac{x_i - \min(X)}{\max(X) - \min(X)} \tag{1}$$

Here, $\min(X)$ represents the minimum value in the feature vector, and $\max(X)$ denotes the maximum value. By subtracting the minimum value and dividing by the range of the data, min–max normalization brings the feature values within the desired range i.e., [0–1]. This normalization technique is particularly useful when the range of the input data is known or when comparing values across different features. Min–max normalization ensures that all features have equal importance during model training and prevents any single feature from dominating the learning process.

3.4. Correlation Analysis

To enhance the forecasting performance of our data-driven model, it is crucial to select appropriate input feature variables. To achieve this, we conducted correlation analysis among the available data sets [33]. The datasets considered six weather forecast features, which include Total Solar Irradiance (TSI), Direct Normal Irradiance (DNI), Global Horizontal Irradiance (GHI), Air Temperature (AT), Pressure, and Relative Humidity (RH). Among these features, total solar irradiance exhibited the highest Pearson correlation coefficient (PCC) with the power output, as shown in Fig. 5 and 6. This finding indicates a strong relationship between the total solar irradiance and the generated power, highlighting the importance of this variable for accurate forecasting. The statistical analysis is also performed on the datasets and the Min, Max, Mean and standard deviation values of the features and output is presented in Table 2.

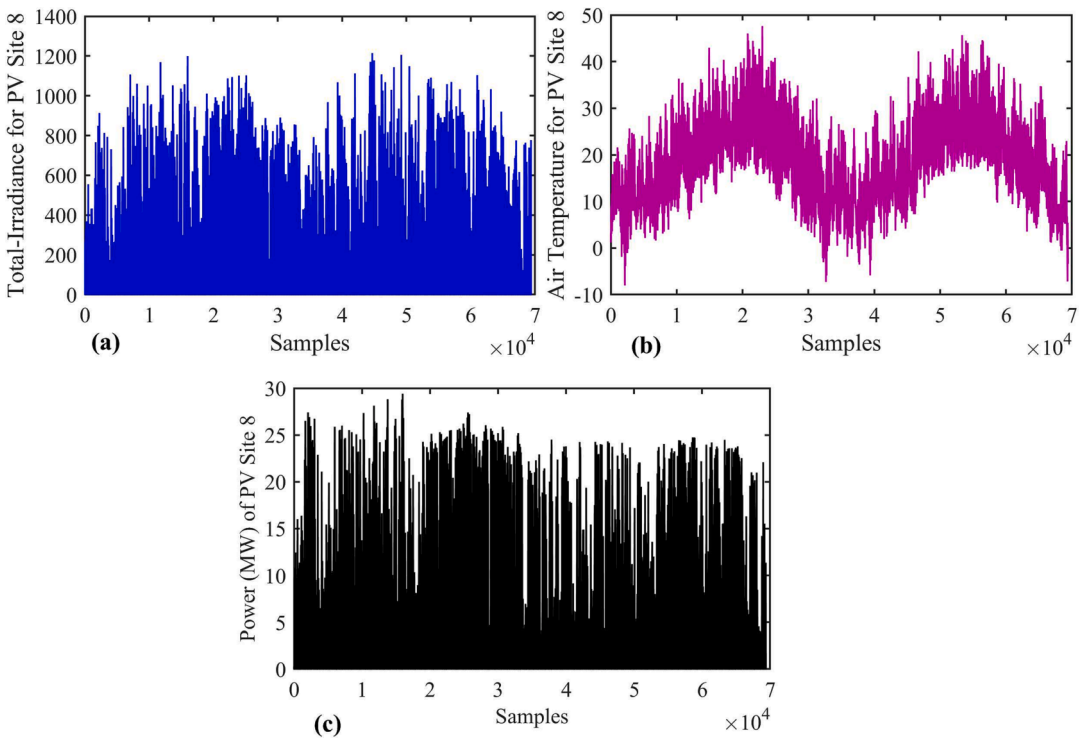

Fig. 4. Few samples from the available datasets from selected solar sites.

3.5. Objective Function and Evaluation Criteria's

The suggested approach is trained and validated using an objective function (fitness function). Lesser values of the utility function illustrate how closely the model's data projections match reality. As a result, the fitness function determines the prediction accuracy.The MSE is the most used fitness function [34] and can be expressed as follows:

$$F.F = \frac{1}{n}\sum_{i=1}^{n}(T_i - P_i)^2 \tag{2}$$

where T_i,P_i, and n represent the true values, anticipated values, and the tally of samples, respectively.

The evaluation of alternative models also employs a number of error indexes. NMSE may be used to verify the degree of dispersion of the results as shown in Eq. 3. Forecast deviation indicators MAE and RMSE are provided in the equations Eq. 4 and 5, respectively. The correlation between actual and predicted values may also be found using the R^2 value, as indicated in Eq. 6.

$$NMSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

$$MAE = \frac{1}{M}\sum_{a=1}^{M}\left|T_a - P_a\right| \times 100\% \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$R^2 = \frac{\sum_{a=1}^{M}\left(T_a - \overline{T}\right)(P_a - \overline{P})^2}{\sum_{a=1}^{M}\left(T_a - \overline{T}\right)\sum_{a=1}^{M}\left(P_a - \overline{P}\right)} \tag{6}$$

where $\overline{T}$ is the average value of the true output, and $\overline{P}$ represents the average value of the expected output.

4. Proposed Technique

This section elaborates on the power forecasting techniques presented in this research work. A total of three techniques are proposed which include a combination of Kepler Optimization Algorithm (KOA) for tuning of hyperparameters of DLMs, Convolution Neural Networks (CNN) based DLM, and Bi-LSTM which is a type of Recurrent Neural Network (RNN).

4.1. Kepler Optimization Algorithm (KOA)

The KOA is a novel optimization method inspired by Kepler's laws of planetary motion [35]. It visualizes the search space by considering the central star as the focal point. In the Keplerian-inspired Optimization Algorithm (KOA), the objects represent candidate solutions and occupy distinct positions relative to the central star at different times. This approach enables efficient exploration and exploitation of the search space. Fig. 7 provides insights into the relationship between an object's position, mass, attractive force from the central star, and orbital velocity. These factors collectively determine the object's proximity to the best solution, represented by the central star. Additionally, the figure illustrates the predominant clockwise rotation of objects, shedding light on the evolution of the searcher's position in a three-dimensional context.

4.1.1. Initialization

In this particular procedure, a collection of planets, comprising N units and referred to as the population size, is dispersed randomly across d dimensions. The arrangement of the planets within this space conforms to the subsequent formula:

$$X_i^j = X_{i,\text{low}}^j + \text{rand}_{[0,1]} \times \left(X_{i,\text{up}}^j - X_{i,\text{low}}^j\right), \begin{cases} i = 1, 2, \ldots, N. \\ j = 1, 2, \ldots, d. \end{cases} \tag{7}$$

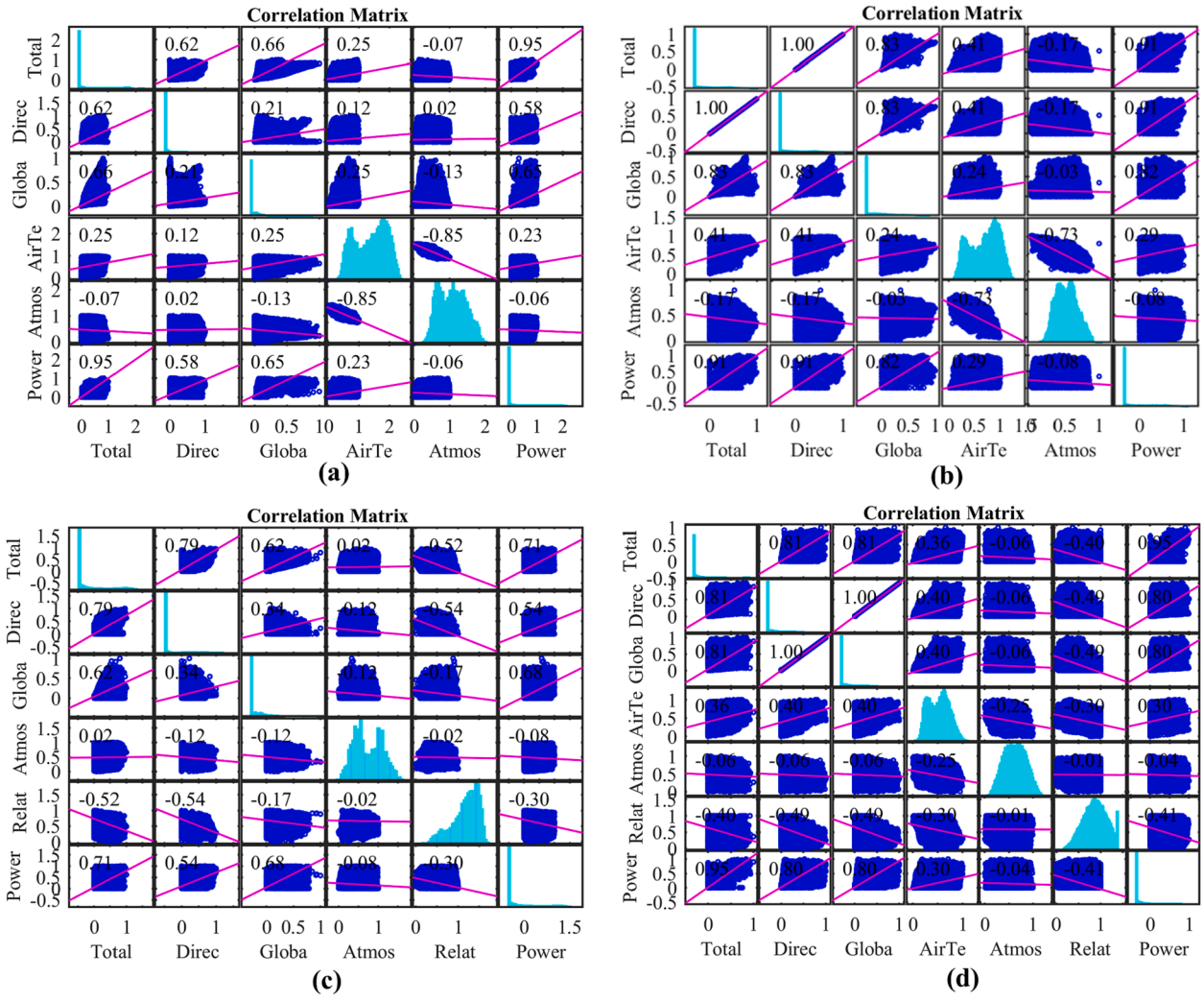

Fig. 5. Correlation matrix of the first 4 PV sites.

In this equation, X_i^j represents the i^{th} celestial body functioning as a potential solution within the exploration space. N denotes the total count of solution candidates available in the exploration space. The variable d corresponds to the dimensionality of the specific optimization problem. $X_{i,\text{up}}^j$ and $X_{i,\text{low}}^j$ signify the upper and lower limits, respectively, for the i^{th} decision variable. Furthermore, rand[0,1] indicates a randomly generated number within the range of 0 to 1. In order to initialize the orbital eccentricity e for each object indexed by i, Eq. 8 is utilized:

$$e_i = \text{rand}_{[0,1]}, i = 1, \ldots, N \tag{8}$$

where, rand[0,1] represents a randomly generated value within the interval [0,1]. Lastly, to initialize the orbital period (T) for each i^{th} object, Eq. 9 is employed:

$$T_i = |r|, i = 1, \ldots, N \tag{9}$$

In this equation, r represents a randomly generated number based on a normal distribution.

4.1.2. Gravitational Force (F)

The attractive force between the Sun X_s and a planet X_i follows the universal law of gravitation, which can be mathematically expressed as:

$$\boldsymbol{F}_{g_i}(t) = e_i \times \mu(t) \times \frac{\overline{M}_s \times \overline{m}_i}{\overline{R}_i^2 + \varepsilon} + r_1 \tag{10}$$

where the normalized masses of the central star X_s (represented by $\widetilde{M_s}$) and the planet X_i (represented by $\widetilde{m_i}$) are taken into account. The variables ∈ (small value) and μ (universal gravitational constant) play important roles in the algorithm. The eccentricity of a planet's orbit, denoted as e_i, introduces a stochastic element to the KOA method with a value ranging between 0 and 1. The term r_1 adds variability to the gravitational values as a randomly generated value between 0 and 1. Lastly, R_i represents the normalized Euclidean distance between X_s and X_i in the optimization process.

$$R_i(t) = \|X_S(t) - X_i(t)\|_2 = \sqrt{\sum_{j=1}^{d} \left(X_{Sj}(t) - X_{ij}(t)\right)^2}. \tag{11}$$

The expression $|X_s(t) - X_i(t)|$ denotes the Euclidean distance between the dimensions of X_s and X_i. To calculate the mass of the Sun and the object i at time t, a straightforward method involves utilizing the fitness evaluation, particularly in the context of a minimization problem. This can be expressed as follows:

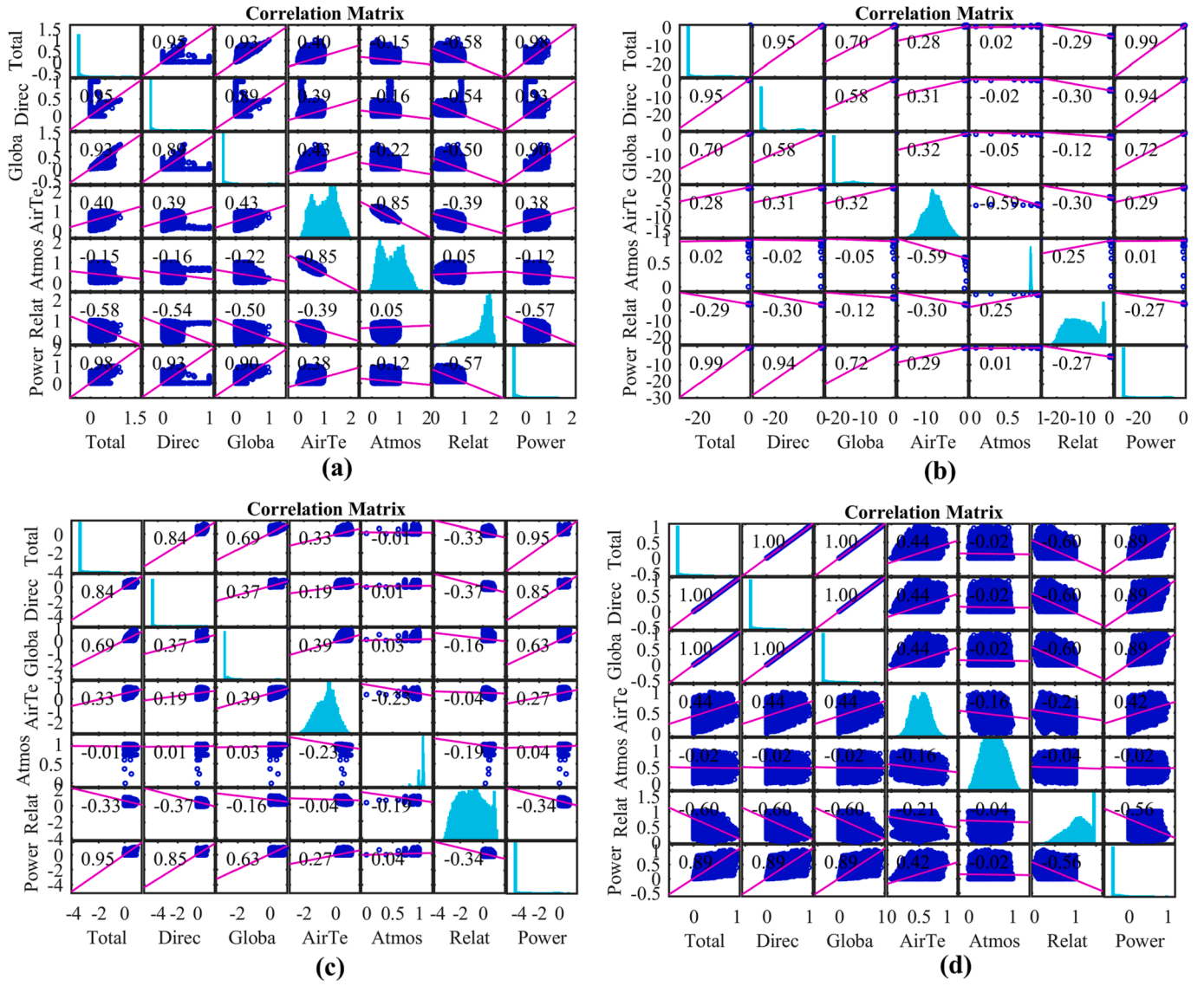

Fig. 6. Correlation matrix of the last 4 PV sites.

$$M_s = r_2 \frac{fit_s(t) - w\text{orst}(t)}{\sum_{k=1}^{N} (fit_k(t) - w\text{wrst}(t))}, \tag{12}$$

$$m_i = \frac{fit_i(t) - \text{worst}(t)}{\sum_{k=1}^{N} (fit_k(t) - \text{worst}(t))}, \tag{13}$$

where

$$fit_s(t) = \text{best}(t) = \min_{k \in 1,2,\ldots,N} fit_k(t), \tag{14}$$

$$\text{worst}(t) = \max_{k \in 1,2,\ldots,N} fit_k(t), \tag{15}$$

In the given equation, r_2 represents a randomly generated number between 0 and 1, serving to introduce divergence in the mass values across different planets. The function μ(t) is responsible for exponentially decreasing with time (t), thus regulating the search accuracy. It can be defined as follows:

$$\mu(t) = \mu_0 \times \exp\left(-\gamma \frac{t}{T_{\max}}\right), \tag{16}$$

where, the symbol γ represents a constant value, μ_o denotes an initial value, and t and T_{max} refer to the current iteration number and the maximum number of iterations, respectively.

4.1.3. Object's Velocity Calculation

The velocity of a celestial object is intricately linked to its position relative to the central star. When a planet is near the star, its velocity increases, and as it moves farther away, its velocity decreases. This modulation in velocity is directly influenced by the gravitational force exerted by the star. When the object is in close proximity to the star, the intensified gravitational pull causes acceleration, resisting a closer approach. Conversely, when the object is distant, the weakened influence of the star's gravity leads to deceleration. Mathematically, these relationships are captured by Eq. 17 using the vis-viva equation, enabling velocity calculations for objects orbiting the star. The equation has two parts: one calculating velocities for planets near the central star, considering distance and fostering diverse search strategies in the KOA approach; the other focuses on planets far from the star, aiming to reduce their velocities compared to the first part. Both parts include a

step size to maintain velocity diversity throughout the optimization process and overcome potential limitations related to reduced diversity and escaping local optima.

$$V_i(t) = \begin{cases} \ell \times \left(2r_4\overrightarrow{X}_i - \overrightarrow{X}_b\right) + \ddot{\ell} \times \left(\overrightarrow{X}_a - \overrightarrow{X}_b\right) + (1 - R_{i-norm}(t)) \\ \times \mathscr{F} \times \overrightarrow{U}_1 \times \overrightarrow{r}_5 \times \left(\overrightarrow{X}_{i,up} - \overrightarrow{X}_{i,low}\right), \text{if } R_{i-norm}\left(t\right) \leqslant 0.5 \\ r_4 \times \mathscr{L} \times \left(\overrightarrow{X}_a - \overrightarrow{X}_i\right) + (1 - R_{i-norm}(t)) \\ \times \mathscr{F} \times U_2 \times \overrightarrow{r}_5 \times \left(r_3\overrightarrow{X}_{i,up} - \overrightarrow{X}_{i,low}\right), \text{Else} \end{cases} \tag{17}$$

$$\ell = \overrightarrow{U} \times \mathscr{M} \times \mathscr{L} \tag{18}$$

$$\mathscr{L} = \left[\mu\left(t\right) \times (M_S + m_i)\left|\frac{2}{R_i(t) + \varepsilon} - \frac{1}{a_{i(t)} + \varepsilon}\right|\right]^{\frac{1}{2}} \tag{19}$$

$$\mathscr{M} = (r_3 \times (1 - r_4) + r_4\,), \tag{20}$$

$$\overrightarrow{U} = \begin{cases} 0 & \overrightarrow{r_5} \leqslant \overrightarrow{r_6} \\ 1 & \text{Else} \end{cases} \tag{21}$$

$$\ddot{\ell} = \left(1 - \overrightarrow{U}\right) \times \overrightarrow{\mathscr{M}} \times \mathscr{L}, \tag{22}$$

$$\overrightarrow{\mathscr{M}} = (r_3 \times (1 - \overrightarrow{r_5}) + \overrightarrow{r_5}\,), \tag{23}$$

$$\overrightarrow{U}_1 = \begin{cases} 0 & \overrightarrow{r_5} \leqslant r_4 \\ 1 & \text{Else} \end{cases} \tag{24}$$

$$U_2 = \begin{cases} 0 & r_3 \leqslant r_4 \\ 1 & \text{Else} \end{cases} \tag{25}$$

where, $\overrightarrow{V_i}$ represents the velocity of object i at time t, while X_i refers to the object itself. The variables r_3 and r_4 are random numeric values generated within the range of [0, 1]. The vectors $\overrightarrow{r_5}$ and $\overrightarrow{r_6}$ contain random values between 0 and 1. The solutions X_a and X_b are selected randomly from the population. The masses of X_s and X_i are denoted as M_s and m_i, respectively. The term μ(t) represents the universal gravitational constant, and ϵ is a small value introduced to avoid division by zero errors. The distance R_i(t) corresponds to the distance between the best solution X_s and the object X_i at time t. Lastly, a_i denotes the semi-major axis of the elliptical orbit of object i at time t, which is determined using Kepler's third law. This can be mathematically expressed as follows:

$$a_i\left(t\right) = r_3 \times \left[T_i^2 \times \frac{\mu(t) \times (M_s + m_i)}{4\pi^2}\right]^{\frac{1}{3}} \tag{26}$$

where the orbital period of object i, denoted as T_i. In this algorithm, the semi-major axis of object i elliptical orbit gradually decreases with successive generations. This approach aims to guide the solutions towards the global best solution. *R_i-norm (t)* represents the normalized Euclidean distance between X_s and X_s at time t. It can be computed as follows:

$$R_{i-\text{norm}}\left(t\right) = \frac{R_i(t) - \min(R(t))}{\max(\text{R(t)}) - \min(R(t))} \tag{27}$$

When *R_i-norm (t)* is less than or equal to 0.5, it indicates that the object is in close proximity to the Sun. In such cases, the object will increase its speed to counteract the strong gravitational force exerted by the Sun and avoid drifting towards it. Conversely, if *R_i-norm (t)* is greater than 0.5, the object will slow down its speed. This adjustment helps maintain stability and prevents excessive acceleration as the object moves further away from the Sun.

4.1.4. Local Optimum Escaping

Within the solar system, the majority of objects follow a

Table 2
Statistics of the datasets of PV sites.

PV Site	Stats	TSI (W/m^2)	DNI (W/m^2)	GHI (W/m^2)	AT (deg$C$)	Pressure (hpa)	RH (%)	Power (MW)
Site 1	Max	1359	980	989	41.2	936.3	-	48.3
	Min	0	0	0	−18.2	894	-	0
	Mean	266.20	93.25	67.69	13.14	913.36	-	9.66
	Std	367.89	200.77	111.19	14.33	8.74	-	13.70
Site 2	Max	1041.93	751.75	561.80	40.47	881.67	-	109.36
	Min	0	0	0	−13.92	844.51	-	.
	Mean	169.30	122.15	78.29	13.69	861.03	-	19.56
	Std	248.07	178.98	117.58	12.03	6.147	-	27.93
Site 3	Max	1117	760	656	1038.60	80.50	-	29.91
	Min	0	0	0	994.80	14.10	-	−0.063
	Mean	198.81	100.72	69.30	1016.01	58.24	-	5.44
	Std	294.57	185.09	101.87	9.323	13.15	-	8.25
Site 4	Max	1237.40	1010.27	150.96	49.79	1100.31	100	114.68
	Min	0	0	0	−5.318	928.59	18.50	−0.440
	Mean	150.15	139.51	20.84	18.71	1011.37	66.24	21.44
	Std	253.43	210.68	31.48	10.27	33.22	17.23	27.11
Site 5	Max	1467	1962	1208	39.50	1039.40	93.20	99.55
	Min	0	0	0	−6.6	990.7	10.60	−0.540
	Mean	164.59	148.10	115.27	17.78	1011.99	71.58	14.51
	Std	273.74	235.13	203.41	9.63	9.94	15.62	23.88
Site 6	Max	1365.40	1179.80	296.20	36.69	846.07	97.90	31.23
	Min	0	0	0	2.946	389.82	1.413	0
	Mean	243.08	215.14	53.92	20.62	830.67	53.94	6.364
	Std	355.43	337.60	69.358	5.756	4.617	23.85	9.166
Site 7	Max	1393.73	1095.40	1125.13	30.12	867.10	100.94	29.77
	Min	0	0	0	−4.263	398.20	1.9533	0
	Mean	206.08	182.99	108.68	13.739	842.93	55.180	5.409
	Std	299.91	306.83	190.63	5.9266	24.397	23.917	8.0422
Site 8	Max	1214.54	1056.6	157.89	47.630	1037.78	100	29.41
	Min	0	0	0	−8.040	881.40	11.83	0
	Mean	163.24	142.02	21.221	18.009	956.41	71.70	4.23
	Std	245.39	213.49	31.901	8.5622	30.535	18.50	6.51

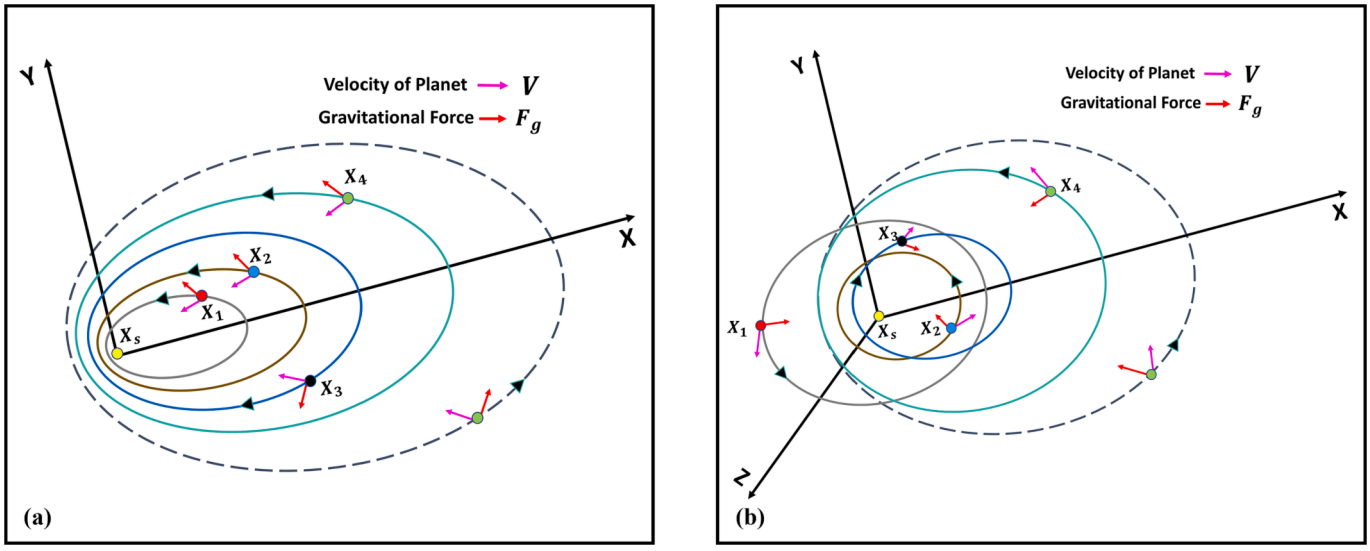

Fig. 7. Working Principle of KOA algorithm.

counterclockwise orbital path around the Sun, accompanied by their own rotational motion. However, there exist certain objects that exhibit a distinct behavior by orbiting around Sun in a clockwise direction. Exploiting this characteristic, the proposed KOA algorithm leverages the phenomenon to effectively break free from local optimum regions. This is achieved through the utilization of a flag, denoted as F, which dynamically alters the search direction. By strategically adjusting the search direction, agents within the algorithm gain enhanced capabilities to meticulously explore the search space, increasing the likelihood of accurate and comprehensive scanning.

4.1.5. Objects' Positions Updation

Fig. 8 provides a visual depiction of the separate exploration and exploitation regions surrounding the central star. Now, let us delve into the specific particulars of the exploration and exploitation phases. During the exploration phase, the algorithm focuses on celestial objects that are distant from the central star, thereby improving its capacity to thoroughly explore the entire search space. Consistent with the preceding steps, the updated position of each object located far from the star is computed using Eq. 28:

$$\vec{X}_i\left(t+1\right) = \vec{X}_i\left(t\right) + \mathscr{F} \times \vec{V}_i\left(t\right) + \left(\boldsymbol{F}_{g_i}(t) + |r|\right) \times \vec{U} \times \left(\vec{X}_S\left(t\right) - \vec{X}_i\left(t\right)\right), \tag{28}$$

$$\vec{X}_i\left(t+1\right) = \vec{X}_i\left(t\right) \times \vec{U}_1 + \left(1 - \vec{U}_1\right) \times \left(\vec{X}_i\left(t\right) + \frac{\vec{X}_S + \vec{X}_a\left(t\right)}{3.0} + h \times \left(\vec{X}_i\left(t\right) + \vec{X}_S + \frac{\vec{X_a}\left(t\right)}{3.0} - \vec{X}_b\left(t\right)\right)\right) \tag{29}$$

Within the equation, $\vec{X}_i\ (t+1)$ indicates the revised position of object i at time $(t+1)$, $\vec{V}_i\ (t)$ represents the velocity required for object i to reach its new position, $\vec{X}S(t)$ denotes the best position of the central star discovered thus far, and F_{gi} is a flag that guides the search direction.

4.1.6. Central Star Distance Updation

The value of the regulating parameter h plays a crucial role in determining the balance between exploration and exploitation. When h is set to a high value, the emphasis is on the exploration operator, leading to an expansion of the orbital separation between the celestial objects and the central star. Conversely, when h assumes a low value, the exploitation operator takes priority, enabling focused exploration in the proximity of the best solution obtained so far, especially when the distance between the central star and the celestial objects is minimal.

In order to enhance the exploration and exploitation operators within the KOA framework, this principle is randomly incorporated into Eq. 28 during the execution of the algorithm. The aim of this integration is to improve the adaptability and effectiveness of the algorithm. Mathematically it can be described as follows:

In this equation, the adaptive factor h determines the distance between the central star (Sun) and the current planet at time t. The value of h is computed using the following procedure:

Fig. 8. Mapping of Exploration and Exploitation Zones centered on the Central Star.

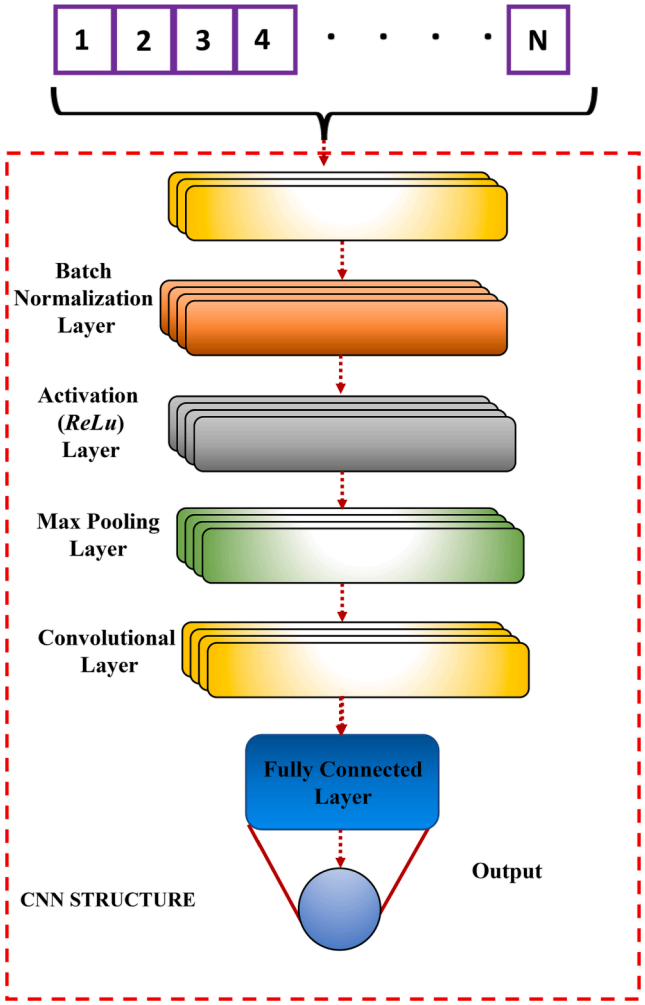

Fig. 9. The structure of an CNN model illustrating different layers.

$$h = \frac{1}{e^{\eta r}} \tag{30}$$

where r is a randomly generated number following a normal distribution, and η is a linearly decreasing factor that ranges from 1 to −2. The value of η is determined as follows:

$$\eta = (a_2 - 1) \times r_4 + 1, \tag{31}$$

In the equation, the cyclic controlling parameter a_2 gradually decreases from −1 to −2 over the course of T cycles throughout the entire optimization process. The value of a_2 is determined according to the following scheme:

$$a_2 = -1 - 1 \times \left(\frac{t \% \frac{T_{\max}}{\overline{T}}}{\frac{T_{\max}}{\overline{T}}} \right) \tag{32}$$

4.1.7. Elitism

This step incorporates an elitist strategy to preserve the best positions for both the planets and the Sun. In this approach, the fitness values of both previously existing and newly generated particles are compared. The particle with the superior accuracy between the two is then selected and kept in the population. The procedure can be summarized using Eq. 33:

$$\overrightarrow{X}_{i,\text{new}}(t+1) = \begin{cases} \overrightarrow{X}_i(t+1), \text{if } f\left(\overrightarrow{X}_i(t+1)\right) \leqslant f\left(\overrightarrow{X}_i(t)\right) \\ \overrightarrow{X}_i(t), Else \end{cases} \tag{33}$$

4.2. Convolutional Neural Network (CNN)

CNNs, or Convolutional Neural Networks, represent a highly effective class of deep learning models extensively applied in computer vision, speech recognition, and natural language processing [36]. They excel in handling spatial and temporal data, making them ideal for tasks like power forecasting in renewable energy applications using time series data. The fundamental concept of CNNs involves utilizing convolutional layers to autonomously learn hierarchical representations from the input data, which are subsequently passed to fully connected layers for classification or regression. The layer structure of a CNN network is depicted in Fig. 9. Additionally, a one-dimensional CNN (1D-CNN) is a specialized version of the standard CNN architecture tailored for processing one-dimensional input data, like time-series data or sequences of feature vectors [37]. In a 1D-CNN, input data undergoes convolution with a set of filters, each sliding over the input in a single dimension. Following this, the output of each filter is activated using a non-linear function like ReLU, and then downsampled using max pooling or average pooling. The resulting feature maps are flattened and inputted into one or more fully connected layers for classification or regression.

The mathematical equations for a 1D-CNN can be expressed as follows. Given an input signal $\mathbf{x} \in \mathbb{R}^{T \times C}$, where T is the length of the signal and C is the number of channels (i.e., the number of features at each time step), we can apply a set of K filters $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_K \in \mathbb{R}^{F \times C}$, where F is the length of each filter. Each filter is convolved with the input signal using valid convolution, resulting in K feature maps $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_K \in \mathbb{R}^{(T-F+1) \times 1}$:

$$\mathbf{z}k = f\left(\mathbf{W}k^* \mathbf{x}\right) = f\left(\sum i = 1^C \mathbf{w}ki^* \mathbf{x}_{:i}\right), \quad k = 1, 2, \ldots, K, \tag{34}$$

where $\mathbf{w}ki$ is the i-th column of the k-th filter, $\mathbf{x} : i$ is the i-th channel of the input signal, and $f(\cdot)$ is the activation function. The output of each filter is then downsampled using max pooling or average pooling,

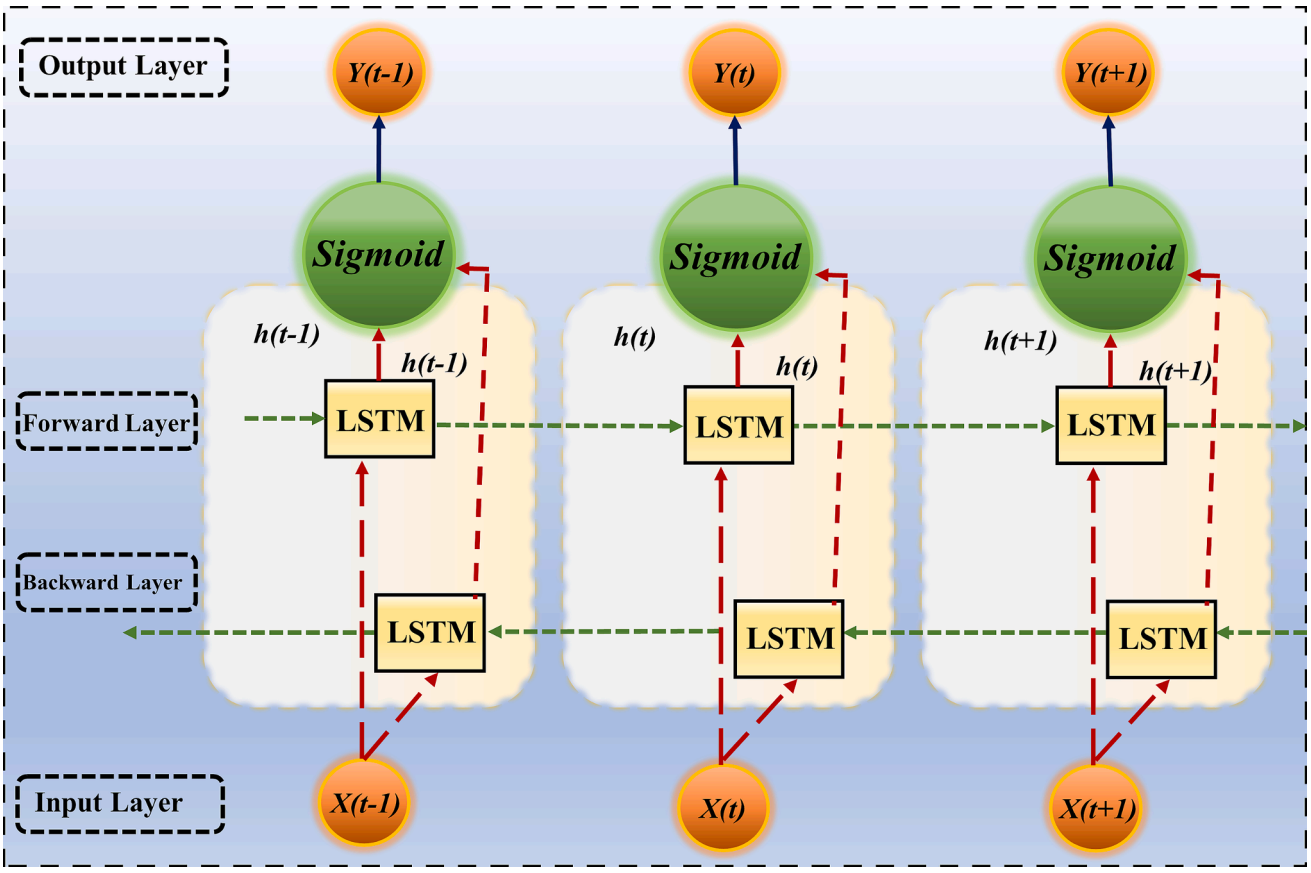

Fig. 10. The structure of an Bi-LSTM cell.

resulting in K pooled feature maps $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_K \in \mathbb{R}^{P\times 1}$:

$$\mathbf{p}_k = \text{pool}(\mathbf{z}_k) = \text{pool}(\mathbf{W}_k * \mathbf{x}), \quad k = 1, 2, \ldots, K, \tag{35}$$

where $\text{pool}(\cdot)$ is the pooling operation, and P is the length of the pooled feature maps. The pooled feature maps are then concatenated into a single feature vector $\mathbf{h}$.

4.3. Bidirectional Long Short-Term Memory Network (Bi-LSTM)

Bidirectional LSTM (Bi-LSTM) is a variant of LSTM that can capture not only past but also future context information by processing the input sequence in both forward and backward directions [38]. The Bi-LSTM cell is made up of two LSTM layers, one of which processes the input sequence in forward direction and the other layer processes in backward direction. The final output at each time step is then produced by concatenating the outputs of the two LSTM layers. The internal structure of Bi-LSTM is shown in Fig. 10.

The forward LSTM layer computes the hidden state $\overrightarrow{\mathbf{h}}_t$ and the memory cell $\overrightarrow{\mathbf{c}}_t$ by processing the input sequence $\mathbf{x}1 : t$ from left to right, while the backward LSTM layer computes the hidden state $\mathbf{h}_t^{\leftarrow}$ and the memory cell $\overleftarrow{\mathbf{c}}\,t$ by processing the input sequence $\mathbf{x}T : t$ from right to left. The final output $\mathbf{h}_t$ at time step t is obtained by concatenating the forward and backward hidden states:

$$\mathbf{h}_t = \left[\overrightarrow{\mathbf{h}}_t; \overleftarrow{\mathbf{h}}_t\right], \tag{36}$$

where $[;]$ denotes concatenation.

The computation of the forward and backward LSTM layers follows the same equations as in the standard LSTM, but with different weight matrices and bias vectors. The equations for the forward LSTM layer are:

$$\overrightarrow{\mathbf{i}}\,t = \sigma\left(\mathbf{W}xi\mathbf{x}t + \mathbf{W}hi\,\overrightarrow{\mathbf{h}}_{t-1} + \mathbf{b}i\right), \tag{37}$$

$$\overrightarrow{\mathbf{f}}\,t = \sigma\left(\mathbf{W}xf\mathbf{x}t + \mathbf{W}hf\,\overrightarrow{\mathbf{h}}\,t-1 + \mathbf{b}f\right), \tag{38}$$

$$\overrightarrow{\mathbf{o}}\,t = \sigma\left(\mathbf{W}xo\mathbf{x}t + \mathbf{W}ho\,\overrightarrow{\mathbf{h}}\,t-1 + \mathbf{b}o\right), \tag{39}$$

$$\overrightarrow{\mathbf{g}}\,t = \tanh\left(\mathbf{W}xg\mathbf{x}t + \mathbf{W}hg\,\overrightarrow{\mathbf{h}}\,t-1 + \mathbf{b}_g\right), \tag{40}$$

$$\overrightarrow{\mathbf{c}}_t = \overrightarrow{\mathbf{f}}\,t \odot \overrightarrow{\mathbf{c}}\,t-1 + \overrightarrow{\mathbf{i}}_t \odot \overrightarrow{\mathbf{g}}_t, \tag{41}$$

$$\overrightarrow{\mathbf{h}}_t = \overrightarrow{\mathbf{o}}_t \odot \tanh\left(\overrightarrow{\mathbf{c}}_t\right). \tag{42}$$

The equations for the backward LSTM layer in the Bi-LSTM cell are similar to those of the forward layer, but with different weight matrices and bias vectors:

$$\overleftarrow{\mathbf{i}}\,t = \sigma\left(\mathbf{W}xi\mathbf{x}t + \mathbf{W}hi\,\overleftarrow{\mathbf{h}}_{t+1} + \mathbf{b}i\right), \tag{43}$$

$$\overleftarrow{\mathbf{f}}\,t = \sigma\left(\mathbf{W}xf\mathbf{x}t + \mathbf{W}hf\,\overleftarrow{\mathbf{h}}\,t+1 + \mathbf{b}f\right), \tag{44}$$

$$\overleftarrow{\mathbf{o}}\,t = \sigma\left(\mathbf{W}xo\mathbf{x}t + \mathbf{W}ho\,\overleftarrow{\mathbf{h}}\,t+1 + \mathbf{b}o\right), \tag{45}$$

$$\overleftarrow{\mathbf{g}}\,t = \tanh\left(\mathbf{W}xg\mathbf{x}t + \mathbf{W}hg\,\overleftarrow{\mathbf{h}}\,t+1 + \mathbf{b}_g\right), \tag{46}$$

$$\overleftarrow{\mathbf{c}}_t = \overleftarrow{\mathbf{f}}\,t \odot \overleftarrow{\mathbf{c}}\,t+1 + \overleftarrow{\mathbf{i}}_t \odot \overleftarrow{\mathbf{g}}_t, \tag{47}$$

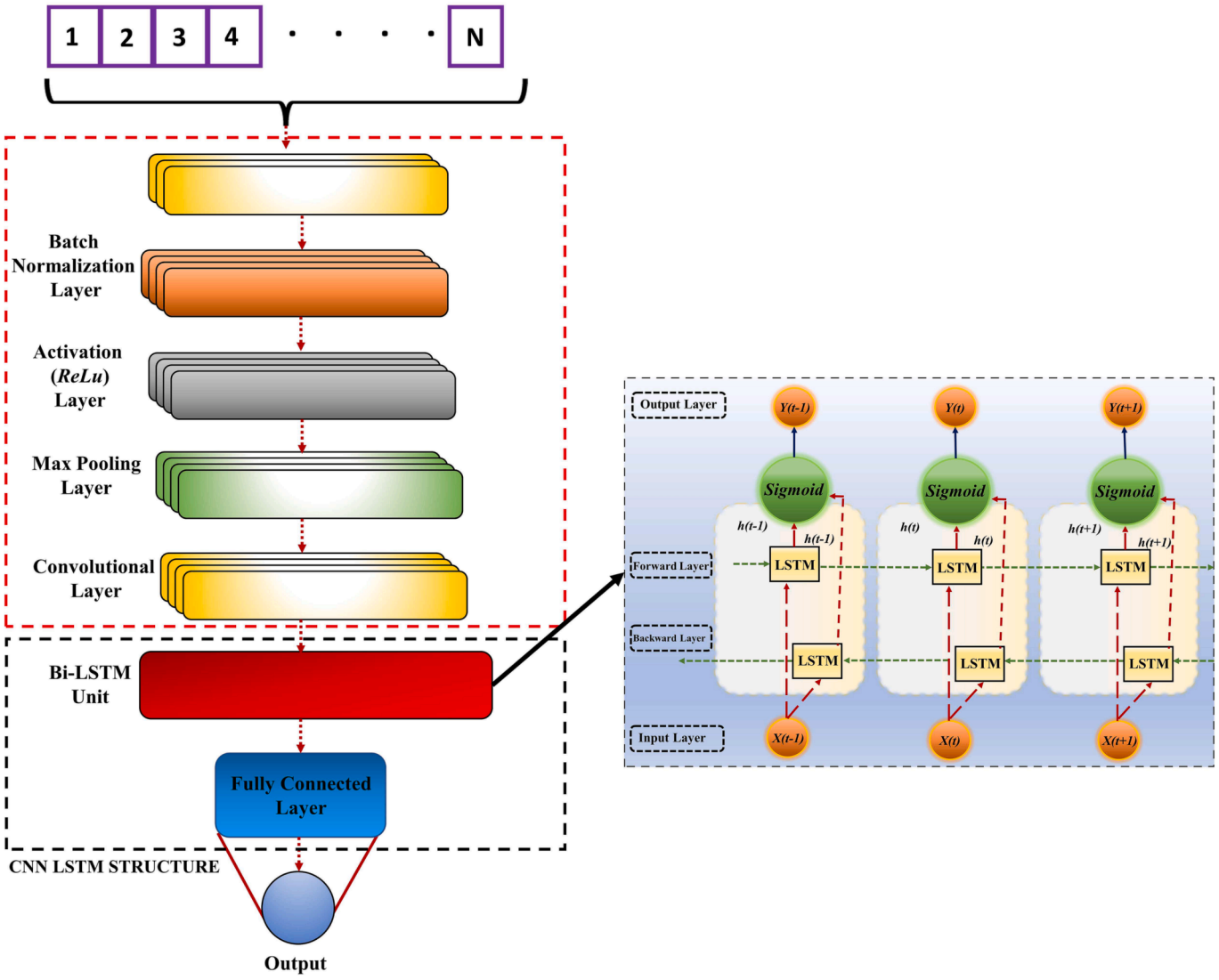

Fig. 11. The structure of an CNN-Bi-LSTM Model.

$$\overleftarrow{\mathbf{h}}_t = \overleftarrow{\mathbf{o}}_t \odot \tanh\left(\overleftarrow{\mathbf{c}}_t\right). \tag{48}$$

The final output at each time step is obtained by concatenating the forward and backward hidden states, $\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t; \overleftarrow{\mathbf{h}}_t]$. The Bi-LSTM cell has been shown to be effective in capturing bidirectional context information and improving the accuracy of tasks such as speech recognition, named entity recognition, and sentiment analysis.

The Bi-LSTM network has several advantages for time series data. By processing the data in both forward and backward directions, it captures bidirectional context information and has a better understanding of the temporal relationships between the input data. This is particularly useful in tasks where the prediction depends on both past and future information, such as speech recognition or machine translation. Additionally, the Bi-LSTM can handle input sequences of variable length, which is a common characteristic of time series data. Its ability to model long-term dependencies and handle sequential data makes it a popular choice for many applications, such as predicting stock prices, detecting anomalies in sensor data, and forecasting energy consumption.

4.4. CNN-Bi-LSTM Model

The CNN-Bi-LSTM network combines the strengths of both the CNN and Bi-LSTM models to process time series data. The CNN layer first extracts feature from the input time series data using convolutional filters [39]. The resulting feature maps are then fed into a Bi-LSTM layer to capture the temporal dependencies in the data. The proposed CNN-Bi-LSTM structure is shown in Fig. 11.

The mathematical equations for the CNN-Bi-LSTM model can be written as follows:

Let $\mathbf{x}_t$ denote the input time series data at time step t, and let $\mathbf{F}$ be the set of filters in the CNN layer. The output feature map $\mathbf{h}_t$ at time step t can be computed as:

$$\mathbf{h}_t = \left[\mathbf{h}_t^{(1)}; \mathbf{h}_t^{(2)}; \ldots; \mathbf{h}_t^{(n)}\right], \tag{49}$$

$$\mathbf{h}_t^{(k)} = \mathrm{ReLU}\left(\mathbf{F}_k * \mathbf{x}_t + \mathbf{b}_k\right), \tag{50}$$

where $*$ denotes the convolution operation, $\mathbf{b}_k$ is the bias term for the $\mathbf{k}_{th}$ filter, and ReLU is the rectified linear unit activation function.

The output of the CNN layer is then fed into the Bi-LSTM layer, which consists of a forward LSTM and a backward LSTM:

$$\overrightarrow{\mathbf{h}}_t = \mathrm{LSTM}\left(\mathbf{h}t, \overrightarrow{\mathbf{h}}t-1\right), \tag{51}$$

$$\overleftarrow{\mathbf{h}}_t = \mathrm{LSTM}\left(\mathbf{h}t, \overleftarrow{\mathbf{h}}\,t+1\right), \tag{52}$$

where LSTM denotes the LSTM cell, and $\overrightarrow{\mathbf{h}}\,t-1$ and $\overleftarrow{\mathbf{h}}\,t+1$ are the previous and next hidden states for the forward and backward LSTMs, respectively.

The final output at each time step is obtained by concatenating the forward and backward hidden states, $\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t; \mathbf{h}_t^{\leftarrow}]$. The output can then be used for various tasks, such as classification or regression.

CNN-Bi-LSTM networks offer additional benefits over Bi-LSTM networks for PV power forecasting. The CNN layer can extract spatial features from weather data, such as cloud coverage and precipitation, which are important factors in determining PV power output. By incorporating this spatial information, the CNN-Bi-LSTM can improve the accuracy of PV power forecasting. Additionally, the CNN layer can reduce the dimensionality of the input data, which can lead to faster training and inference times. Overall, the CNN-Bi-LSTM has shown promising results in PV power forecasting applications, outperforming traditional forecasting methods as well as other deep learning models, such as multi-layer perceptron (MLP) and CNN models. As such, the CNN-Bi-LSTM is a powerful tool for accurately forecasting PV power output, which can help improve the efficiency and reliability of PV systems.

4.5. Hyperparameters of CNN-Bi-LSTM Model

The key hyperparameters to consider when training a CNN-Bi-LSTM model include the following: The number and size of convolutional filters in the CNN layers determine what n-gram features are extracted from the input text. More convolution layers with varying filter sizes allow capturing different granularities of features. The CNN typically uses Rectified Linear Unit (ReLU) activation functions after each convolutional layer to introduce non-linearity. The Bi-LSTM layer has hyperparameters like the number of hidden units, number of layers, and dropout rate. More hidden units allow capturing greater feature complexity. Stacking multiple Bi-LSTM layers enables modeling higher-level abstractions of text. Dropout prevents overfitting by randomly setting inputs to zero during training. Additional important hyperparameters are the batch size, learning rate, and number of training epochs. Larger batch sizes reduce noise but may cause underfitting. The learning rate impacts model convergence. More epochs allow better fitting to the training data but increase risk of overfitting. Careful tuning of these CNN and Bi-LSTM hyperparameters can improve the model's ability to extract informative text features for effective classification while preventing overfitting to the training data. The range of the hyperparameters before applying KOA algorithm is shown in Table 3.

4.6. Proposed KOA based CNN-Bi-LSTM Model

While CNN-Bi-LSTM networks have shown promising results in PV power forecasting, they require careful hyperparameter tuning to achieve optimal performance. The CNN layer, for instance, has a number of hyperparameters that need to be tuned, such as filter size, number of filters, and stride. Similarly, the Bi-LSTM layer has hyperparameters such as the number of LSTM units and the dropout rate. Tuning these hyperparameters can be a time-consuming and resource-intensive process. Additionally, CNN-Bi-LSTM networks are prone to overfitting, especially when trained on small datasets. To address this, regularization techniques such as dropout and L2 regularization can be applied, but these also require additional hyperparameter tuning. Overall, while CNN-Bi-LSTM networks offer improved performance over traditional forecasting methods, their hyperparameter sensitivity and overfitting tendencies can make them challenging to tune and optimize.

Table 3
Range of Hyperparameters of CNN-Bi-LSTM.

Variable	Parameters	Range
Convolutional Layers	Number of Filter	$[2^0–2^9]$
	Size of Filter	[1–7]
	Number of Hidden Nodes	[10–500]
Learning Configuration	Learning Rate	$[10^{-5} – -10^{-1}]$
	Dropout Rate	[0, 0.7]

Metaheuristic algorithms have been shown to be effective for hyperparameter tuning of CNN-Bi-LSTM networks. Metaheuristic algorithms are optimization algorithms that can efficiently search through a large hyperparameter space to find the optimal combination of hyperparameters. Examples of metaheuristic algorithms include Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and Simulated Annealing (SA). These algorithms can help reduce the time and resources required for hyperparameter tuning, while also improving the performance of the CNN-Bi-LSTM model. Additionally, some metaheuristic algorithms, such as the Bayesian optimization algorithm, can also incorporate prior knowledge about the hyperparameters to further improve the optimization process. Overall, metaheuristic algorithms are a promising approach for hyperparameter tuning of CNN-Bi-LSTM networks and can help to further improve their performance in PV power forecasting applications.

KOA has been recently proposed as a novel approach for the hyperparameter tuning of machine learning models, including CNN-Bi-LSTM networks used in PV power forecasting. KOA allows for efficient exploration and exploitation of the hyperparameter space by utilizing the principles of Kepler's laws of planetary motion as discussed in Section 4.1. By incorporating the physical laws of motion, KOA provides a unique and innovative approach to hyperparameter tuning, which has the potential to improve the performance of CNN-Bi-LSTM models in PV power forecasting applications. KOA offers a promising approach to the hyperparameter tuning of CNN-Bi-LSTM models in PV power forecasting. The KOA based CNN-Bi-LSTM model is shown in Fig. 12. After applying KOA algorithm the optimized hyperparameters are presented in Table 4.

4.7. Transductive Transfer Learning (TTL)

TTL has gained significant attention among researchers due to its ability to enhance performance in a target domain by leveraging knowledge learned from a source domain. In TTL, both labeled training sets and unlabeled test sets are used to train a model that can predict labels for the unlabeled test set. TTL aims to extract useful information from the unlabeled instances, bridging the gap between labeled and unlabeled data and improving the model's generalization abilities.

One key advantage of TTL is its ability to overcome data scarcity [40]. In many real-world scenarios, obtaining labeled data is expensive or time-consuming. TTL allows models to leverage a larger pool of available data by utilizing both labeled and unlabeled instances, enhancing their learning capabilities [41]. This approach is particularly valuable when labeled data is limited but unlabeled data is abundant. The most widely known example of transductive learning is domain adaptation [42]. Domain adaptation, a form of transfer learning, focuses on adjusting a model initially trained on a source domain to achieve competent performance within a target domain. The target domain may exhibit divergent data distribution characteristics from the source domain. In the context of machine learning, a domain refers to a specific distribution of data, characterized by features and patterns. Therefore, the current framework of TTL employed in this work is the domain adaptation approach as seen in Fig. 13.

4.8. TTL based KOA-C-Bi-LSTM

The collective knowledge of multiple agents moving in congruence in a predefined search space helps the CNN-Bi-LSTM network find optimal parameters. The sequence map to accomplish the task of finding

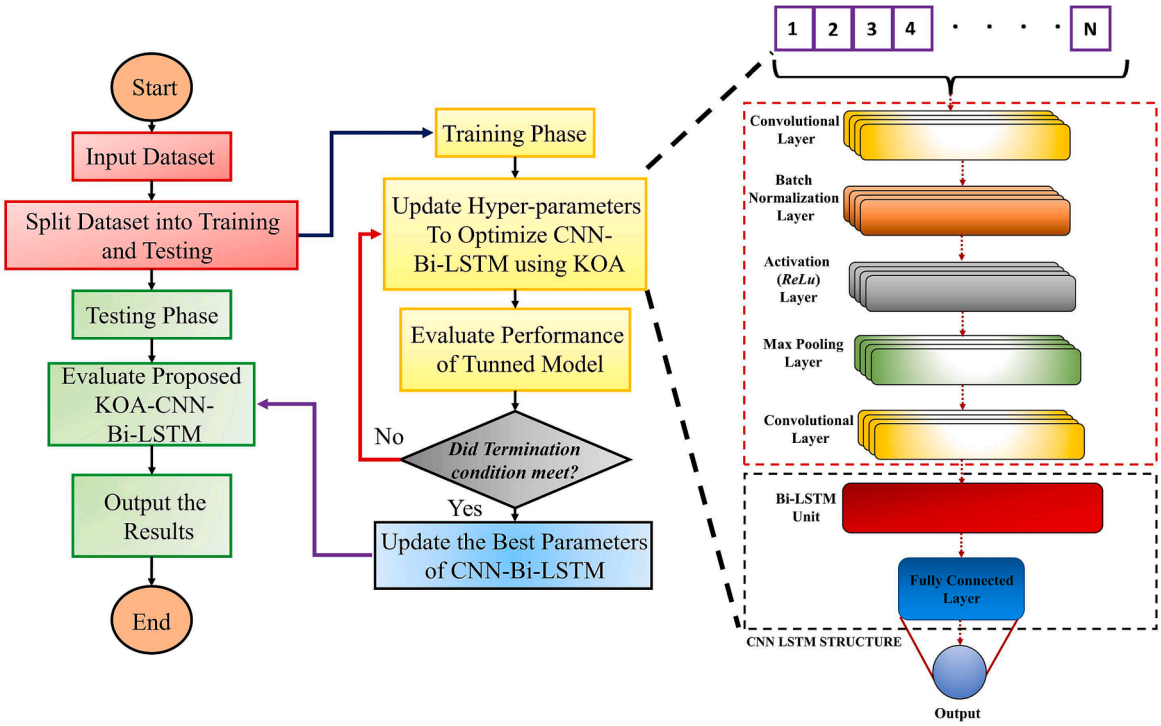

Fig. 12. Flow Chart of KOA based CNN-Bi-LSTM model.

Table 4
Optimized Parameter for CNN Bi-LSTM after COA.

Variable	Parameters	Optimized Value
Convolutional Layers	Number of Filter	64
	Size of Filter	3
	Activation	'ReLU'
Bi-LSTM Layers	Number of Hidden Nodes	100
Learning Configuration	Learning Rate	10^{-2}
	Dropout Rate	0.5

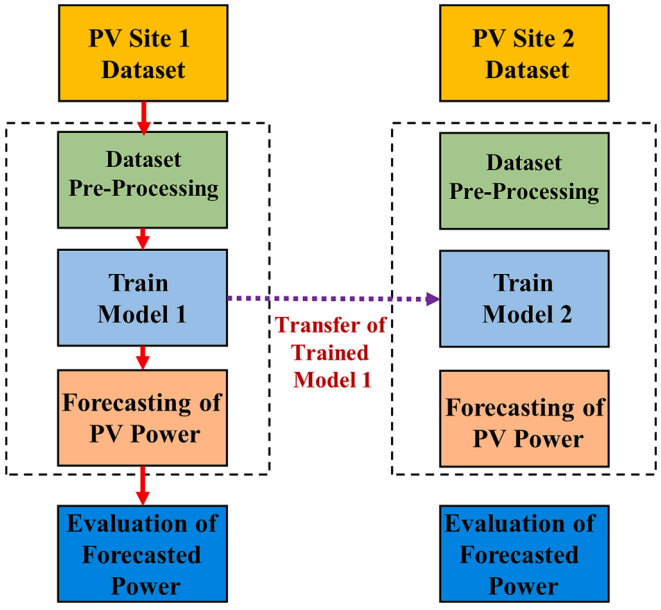

Fig. 13. TTL- Flow structure.

a global best solution in the proposed technique starts with the division of the PV sites datasets into training and testing subsets.

In this research endeavor, we employ a transductive transfer learning strategy to tackle the intricate challenge of analyzing photovoltaic (PV) datasets. We structure our approach into two distinct sets of tasks: source tasks and target tasks. Initially, we curate four source tasks, each associated with a unique PV dataset, representing diverse scenarios. The datasets are meticulously divided into training and testing subsets, with 70% allocated for model training, ensuring that our models are rigorously evaluated for their generalization capability on the remaining 30%. Subsequently, we embark on the core of our strategy–transductive transfer learning. This methodology leverages the knowledge encapsulated within pre-trained Kepler Optimization algoroithm (KOA) based Convolutional Bidirectional Long Short-Term Memory (KOA-C-Bi-LSTM) models obtained from the source tasks. These models serve as invaluable starting points, containing insights into the relationships between PV system performance and environmental factors. The transduction process unfolds in two stages: model initialization and fine-tuning. First, the pre-trained models provide a foundation for the target tasks, offering a preliminary understanding of the underlying PV systems. Then, during the fine-tuning stage, the models are adapted to the intricacies of the target datasets. This step involves updating the model's internal representations and weights while preserving essential domain knowledge extracted from the source tasks. By iteratively applying this transfer learning mechanism to each target task, we aim to enhance the accuracy and predictive power of our TKCBL model for PV dataset analysis. The results section will delve into the evaluation of our models on the target tasks, where we employ rigorous metrics to assess their performance and gauge the effectiveness of knowledge transfer.

5. Results and Discussion

5.1. Models Comparison for PV Power Forecasting

Accurate power forecasting for intelligent power dispatch has a fundamental role in the advancement of smart power grids and the successful implementation of Industry 5.0 as discussed in Section 1 and

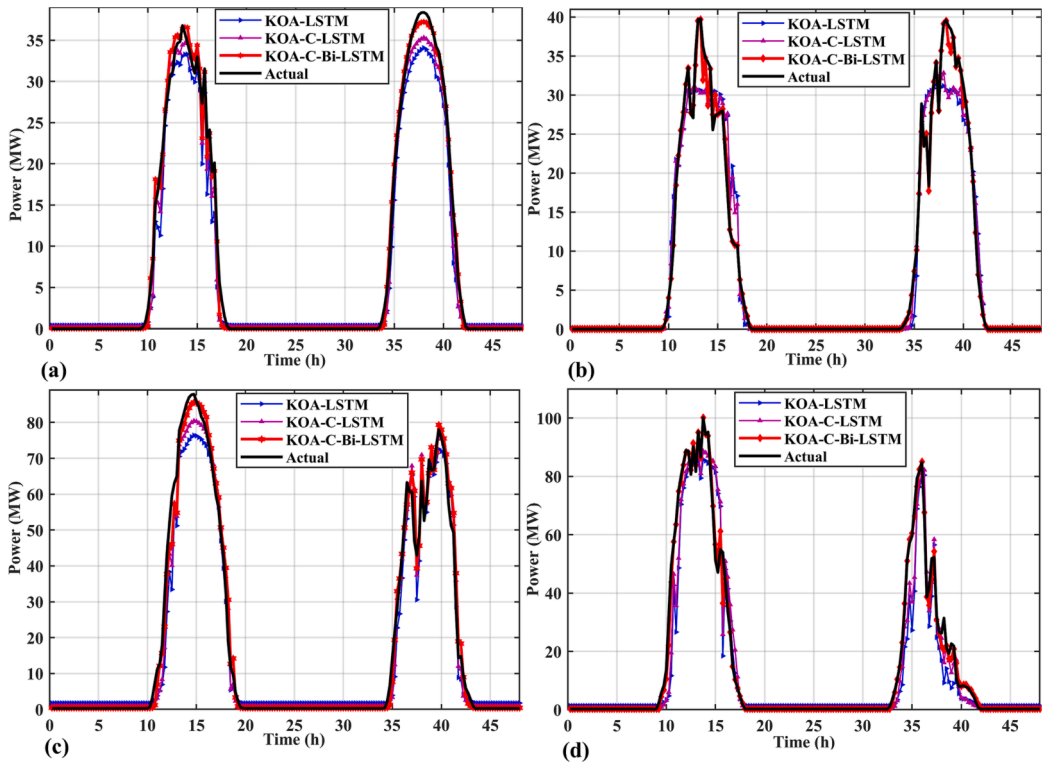

Fig. 14. Power Forecasting Comparison of PV Site 1 and 2 (a) Forecasting in Summer Days of site 1 (b) Forecasting in Winter Days of site 1 (c) Forecasting in Summer Days of site 2 (d) Forecasting in Winter Days of site 2.

Section 2. To identify the most effective technique, three models which are mainly hybrids of KOA, CNN, LSTM, and Bi-LSTM are proposed in this research work. The comparative analysis of different deep learning models trained and tested on various PV sites reveals insightful findings. Three models were evaluated using multiple evaluation metrics including RMSE, MAE, and R-Square. The PV sites power forecasting comparison is shown in Fig. 14–17. The statistical analysis of comparative techniques is presented in Table 3. (See Fig. 18).

5.1.1. PV Sites: Comparison

The comparative analysis of Table 5 and Table 6 provides a comprehensive view of the performance disparities between forecasting techniques applied to solar photovoltaic (PV) energy generation, with Table 5 representing models without Transductive Transfer Learning (TTL) and Table 6 depicting models integrated with TTL. This evaluation serves as a crucial investigation into the potential benefits of TTL for enhancing forecasting precision.

Table 5 reports the results of PV energy generation forecasting for various sites (PV Site 1 to PV Site 4) using three distinct techniques: KOA-C-Bi-LSTM, KOA-C-LSTM, and KOA-LSTM. The key performance metrics, RMSE, NMSE, MAE, and R-Square, reveal distinct characteristics for each combination. The RMSE values in Table 5 range from 0.0017 to 0.02412, indicating variations in forecasting accuracy. Similarly, the NMSE values, spanning from 0.0045 to 0.05401, signify differences in the models' ability to capture the variance in the data, with higher values indicating greater deviations from actual values. The MAE values, ranging from 0.00098 to 0.29404, provide insight into the magnitude of forecasting errors for each technique, while the R-Square values, varying between 91.189 and 98.954, demonstrate how well the models fit the data.

Conversely, Table 6 presents the results of forecasting PV energy generation at different sites (PV Site 5 to PV Site 8) using the same techniques, but this time augmented with TTL, denoted as TKCBL, TKCL, and TKL. The incorporation of TTL in Table 6 leads to noticeable alterations in forecasting accuracy metrics. The RMSE values, now ranging from 0.0012 to 0.08467, indicate potential improvements in forecasting precision when TTL is applied. The NMSE values, spanning from 0.00409 to 0.07364, illustrate the varying degrees to which TTL assists in capturing data variance. Meanwhile, the MAE values, varying between 0.00107 and 0.56102, showcase adjustments in absolute forecasting errors attributable to TTL integration. Finally, the R-Square values, ranging from 93.103 to 99.657, suggest improved model fit due to the introduction of Transductive Transfer Learning.

In comparative terms, it is evident that the introduction of TTL results in considerable enhancements across all forecasting accuracy metrics. TKCBL, TKCL, and TKL consistently outperform their non-TTL counterparts from Table 5, showcasing lower RMSE, NMSE, MAE, and higher R-Square values. This consistent improvement across multiple sites and techniques underscores the pivotal role of TTL in elevating the predictive capabilities of forecasting models in the context of PV energy generation.

The findings from Table 6 emphasize the potential advantages of incorporating Transductive Transfer Learning into renewable energy forecasting practices. These improvements in forecasting accuracy are of paramount importance for efficient energy management and grid integration, as they enable more reliable predictions of solar energy generation. Future research avenues could delve into the specific mechanisms and data characteristics that make TTL particularly effective in this context, as well as explore its broader applicability in enhancing the accuracy of renewable energy forecasting models for various renewable sources and geographical locations.

5.1.2. R-Square Analysis

Certainly, let's focus on a detailed comparison between Table 5 and Table 6 with respect to the R-Square metric, which measures the goodness-of-fit of forecasting models to the data. R-Square values closer to 100 indicate a better fit of the model to the data, suggesting higher predictive accuracy.

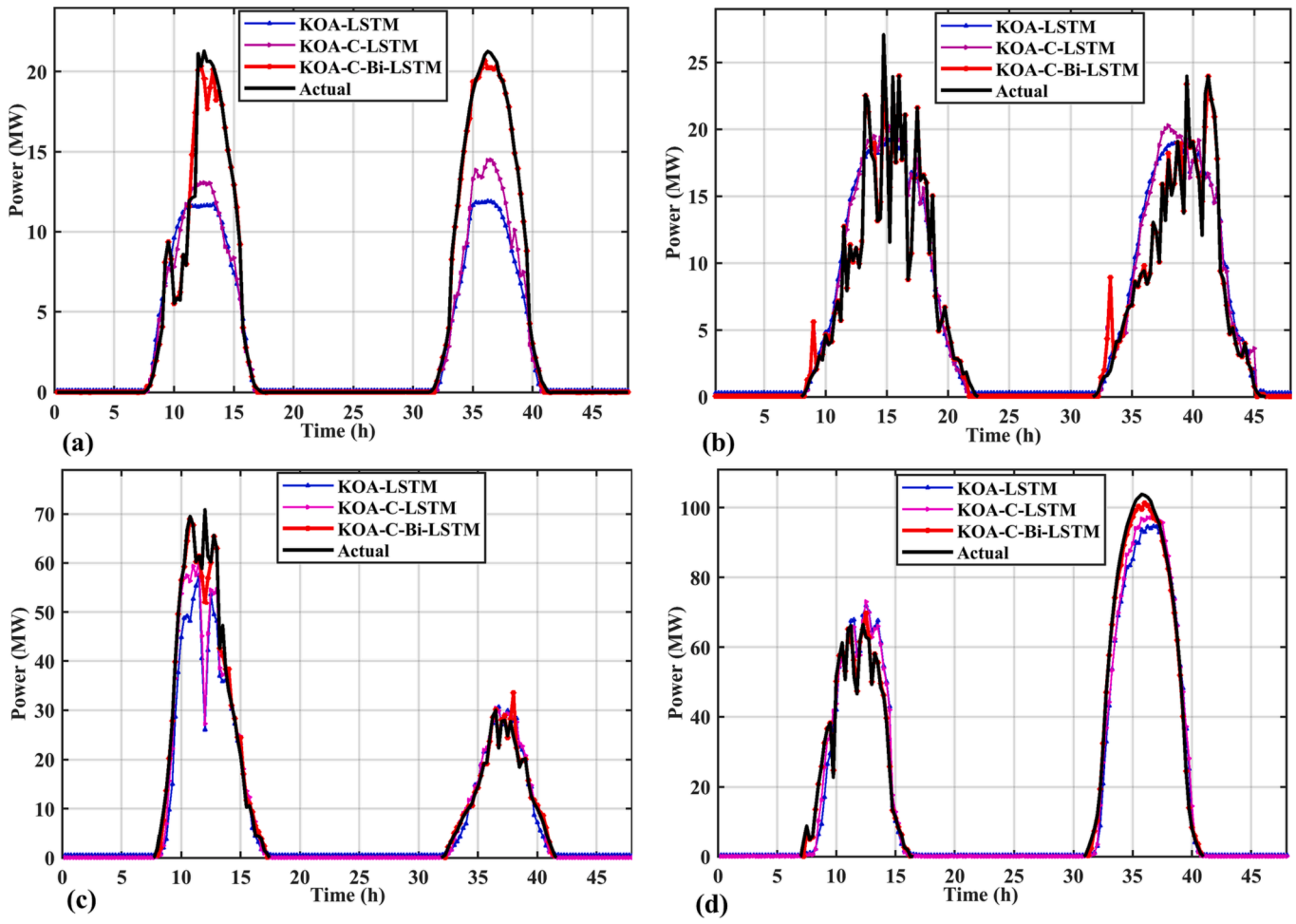

Fig. 15. Power Forecasting Comparison of PV Site 3 and 4 (a) Forecasting in Summer Days of site 3 (b) Forecasting in Winter Days of site 3 (c) Forecasting in Summer Days of site 4 (d) Forecasting in Winter Days of site 4.

In Table 5, which represents forecasting results without Transductive Transfer Learning (TTL), the R-Square values range from 91.189 to 98.954 across different datasets (PV Site 1 to PV Site 4) and techniques (KOA-C-Bi-LSTM, KOA-C-LSTM, and KOA-LSTM). These values indicate that the models in Table 5 exhibit reasonably good fits to the data, with some variations depending on the dataset and technique. For instance, PV Site 1 achieved a high R-Square value of 98.954 when using the KOA-C-Bi-LSTM technique, indicating a strong predictive fit for this specific combination.

In Table 6, which presents forecasting results with the integration of Transductive Transfer Learning (TTL), the R-Square values range from 93.103 to 99.657 for different datasets (PV Site 5 to PV Site 8) and techniques (TKCBL, TKCL, and TKL). Notably, all R-Square values in Table 6 are consistently higher than those in Table 5. This suggests a substantial improvement in model fit and predictive accuracy when TTL is employed. For instance, PV Site 8 achieved an exceptionally high R-Square value of 99.657 when using the TKCBL technique, indicating an exceptionally strong fit of the model to the data.

Comparatively, the R-Square values in Table 6 clearly demonstrate the substantial benefits of incorporating Transductive Transfer Learning into the forecasting techniques. These higher R-Square values indicate that TTL contributes significantly to enhancing the accuracy of the forecasting models. It implies that TTL helps the models better capture the underlying patterns and variability in the data, resulting in improved predictive power.

In summary, the comparison of R-Square values between Table 5 and Table 6 underscores the substantial positive impact of Transductive Transfer Learning on the goodness-of-fit of forecasting models for PV energy generation. These findings suggest that TTL can be a valuable tool in enhancing the accuracy of renewable energy forecasting, which is crucial for efficient energy management and grid integration. Further research can explore the specific mechanisms and circumstances under which TTL is most effective in improving model fit and predictive accuracy.

5.1.3. *Overall Performance*

In an overarching view of Tables 1 and 2, which respectively represent forecasting results without and with Transductive Transfer Learning (TTL), it becomes evident that the integration of TTL has a transformative impact on the overall performance of forecasting models for solar photovoltaic (PV) energy generation. While Table 1 showcases reasonably good predictive accuracy with R-Square values ranging from 91.189 to 98.954, Table 2 reveals a marked improvement across the board, with R-Square values elevated to a consistently higher range of 93.103 to 99.657. This outcome underscores the pivotal role of TTL in elevating the quality of forecasts, indicating a shift from relatively good fits to exceptionally strong fits of the models to the data. These findings are particularly significant in the context of renewable energy management, as they imply that TTL can substantially enhance the precision and reliability of PV energy generation predictions, a critical factor for grid integration and efficient energy utilization.

The performance comparison between Tables 1 and 2 highlights not only the superior predictive capabilities of TTL-enhanced models but also the potential implications for practical applications. The increase in

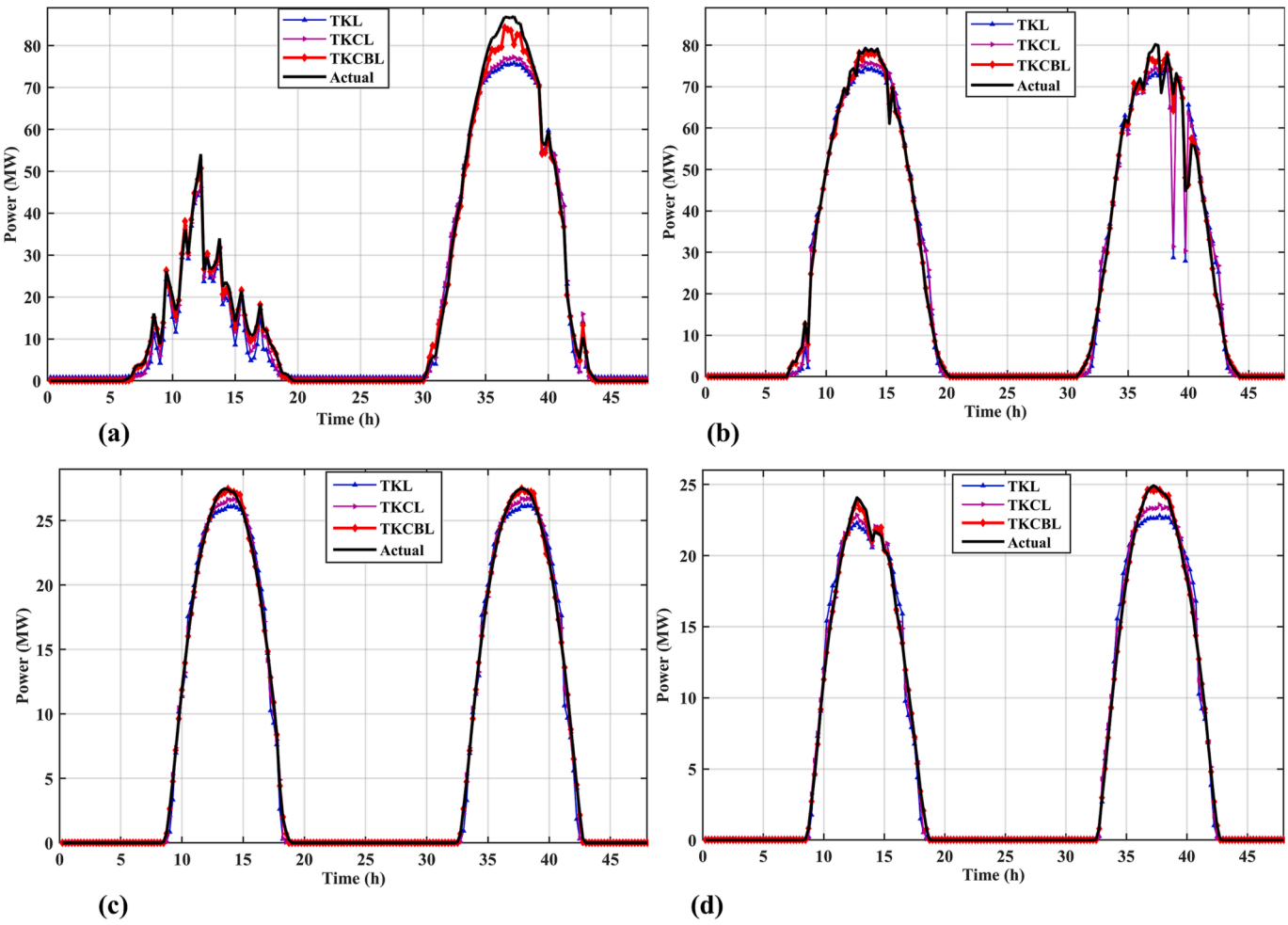

Fig. 16. Power Forecasting Comparison of PV Site 5 and 6 (a) Forecasting in Summer Days of site 5 (b) Forecasting in Winter Days of site 5 (c) Forecasting in Summer Days of site 6 (d) Forecasting in Winter Days of site 6.

R-Square values signifies an enhanced ability to capture and account for the underlying variability in PV energy generation data, which is vital for informed decision-making in renewable energy systems. The findings suggest that the utilization of TTL can significantly contribute to reducing forecasting errors, optimizing energy production, and ultimately advancing the integration of solar energy into the broader energy landscape. Future research endeavors may delve deeper into the mechanisms through which TTL achieves these improvements and explore its adaptability to other renewable energy sources, ultimately paving the way for more accurate and efficient renewable energy forecasting systems.

5.2. Comparative Analysis with Existing Techniques

Correct PV power forecasting is very crucial in integration of large solar power plants with conventional grids. In recent years, several state-of-the-art techniques have been developed for PV site output power prediction. The comparative analysis is presented in Table 7. Among them, HBO-LSTM [43] achieved an R-Square (R2) value of 0.965 with an RMSE of 3.9%. This indicates that the model explains 96.5% of the variance in the target variable but has an average prediction error of 3.9%. Similarly, the ACO-NN technique [44] attained an R-Square value of 0.815 and an RMSE of 3.4%, explaining 81.5% of the variance but with a slightly lower prediction error.

In 2022, the VMD-CNN-BiGRU technique [45] emerged with an RMSE of 2.8%, yet the R-Square value was not provided, making it challenging to assess the model's ability to explain the variance. Furthermore, the DSN technique [46] achieved an RMSE of 2.9%, but the R-Square value was again not reported, limiting a comprehensive analysis of its performance.

In this context, our proposed KOA-C-Bi-LSTM technique, introduced in 2023, showcases remarkable advancements in PV site data prediction. With an R-Square value of 0.993 and an RMSE of 0.31%, the proposed model not only outperforms the previously published techniques but also demonstrates exceptional explanatory power and significantly lower prediction errors. The KOA-C-Bi-LSTM technique exhibits an outstanding ability to explain 99.48% of the variance in the target variable, indicating a high level of accuracy and precision in the predictions.

The performance comparison clearly establishes the superiority of the KOA-C-Bi-LSTM technique in PV site data prediction. Its exceptionally high R-Square value and impressively low RMSE highlight the advancements made compared to the state-of-the-art techniques. Your proposed model achieves a near-perfect explanatory power, capturing almost all the variance in the target variable, while maintaining an astoundingly low average prediction error of only 0.27%. These results demonstrate the efficacy and innovation of the KOA-C-Bi-LSTM technique in improving the accuracy and reliability of PV site data predictions, making it a promising choice for future research and practical applications in the field.

5.3. Granger Causality Test (GCT)

A statistical method called the Granger causality test is used to

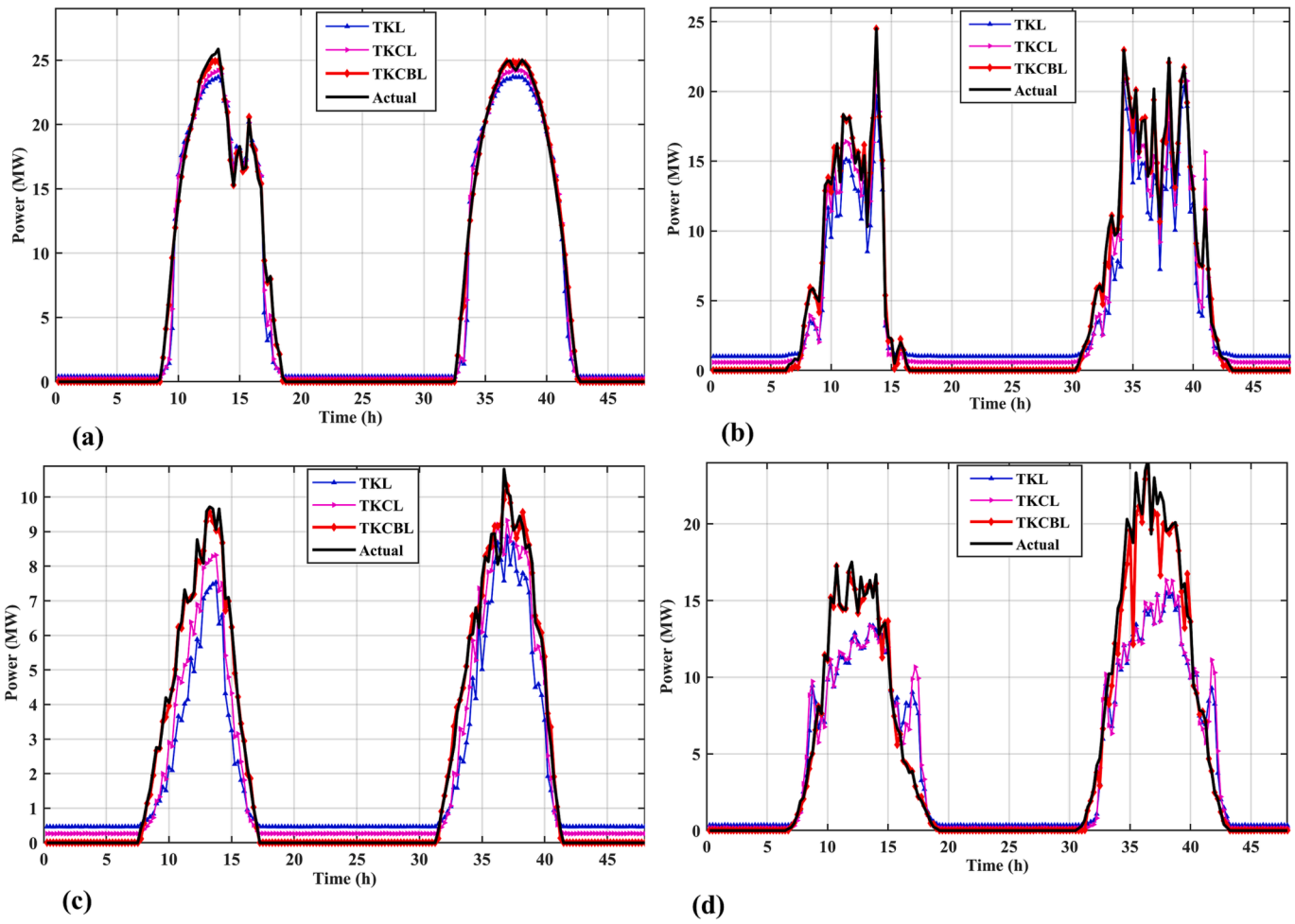

Fig. 17. Power Forecasting Comparison of PV Site 7 and 8 (a) Forecasting in Summer Days of site 7 (b) Forecasting in Winter Days of site 7 (c) Forecasting in Summer Days of site 8 (d) Forecasting in Winter Days of site 8.

examine time series data in order to ascertain if previous values of one variable may predict future values of another variable. Given that it takes into account the delays between solar irradiance and PV power generation, this test can be helpful for projecting short-term photovoltaic (PV) power. The time lag takes into account the possibility that changes in irradiance may not have an immediate effect on PV power. The test offers insights into the dynamics of the PV system by examining these time delays. The link between input variables like irradiance and the output PV power may thus be better captured by including this information into forecasting models.

The Granger causality test [47] evaluates whether changes in a set of "cause" time series variables have altered the probability distribution of a set of "effect" time series variables. In order to prevent autocorrelation, which might skew the results, the data should be preprocessed. The data is further examined for unit roots, which might render the test incorrect. The null and alternative hypotheses, such as the one that variable $y(t)$ does not Granger-cause variable $x(t)$, are stated at the beginning of the study. The test can be repeated with various lag lengths to ensure the reliability of the results. The results shouldn't vary depending on the chosen lag.

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + c_1 + v_1(t) \tag{53}$$

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + \sum_{i=1}^{\infty} \beta_i y(t-j) + c_2 + v_2(t) \tag{54}$$

To check if $y(t)$ Granger-causes $x(t)$, we can consider the following equations:

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + c_1 + u_1(t) \tag{55}$$

$$y(t) = \sum_{i=1}^{\infty} \alpha_i y(t-i) + \sum_{i=1}^{\infty} \beta_i y(t-j) + c_2 + u_2(t) \tag{56}$$

Where $y(t)$ represents the sum of infinite lagged terms $\alpha_i y(t-i)$, along with a constant term c_1 and an error term $u_1(t)$. Similarly, Eq. 56 includes lagged terms $\alpha_i y(t-i)$ and $\beta_i y(t-j)$, where j represents the delays. It also includes a constant term c_2 and an error term $u_2(t)$.

A hypothesis test can be used to evaluate if the independent variables have a statistically significant causal influence on the dependent variable once the models have been specified. Here, the complete model including the independent variables is contrasted with a null model devoid of the independent variables. According to the null hypothesis, changes in the dependent variable are not caused by the independent variables. We can reject the null hypothesis if the entire model considerably outperforms the null model in predicting the dependent variable. Based on the Granger causality theory, this would show that the independent variables do in fact have a causal impact on the dependent variable. The null hypothesis can be disproved by quantifying the statistical significance of the improvement.

The basic form of the Granger causality test is:

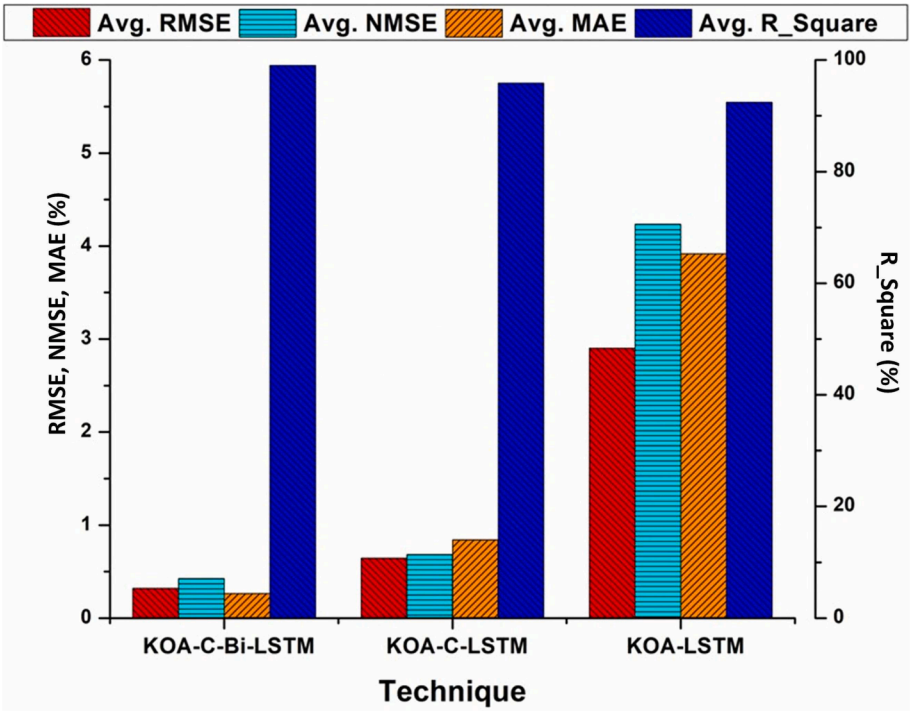

Fig. 18. Statistical analysis average value comparison with competing techniques (a) RMSE (b) NMSE (c) MAE (d) R-Square.

Table 5
Comparison of Statistical Analysis of Forecasting Results with Transductive Transfer Learning.

Dataset	Tech	RMSE	NMSE	MAE	R-Square
PV Site 1	KOA-C-Bi-LSTM	0.0054	0.0018	0.00280	98.954
	KOA-C-LSTM	0.00912	0.00523	0.04841	95.117
	KOA-LSTM	0.01419	0.02011	0.29404	92.487
PV Site 2	KOA-C-Bi-LSTM	0.0026	0.0037	0.00098	99.032
	KOA-C-LSTM	0.00618	0.00469	0.06384	96.531
	KOA-LSTM	0.02412	0.05401	0.19186	91.189
PV Site 3	KOA-C-Bi-LSTM	0.00082	0.00064	0.00120	98.519
	KOA-C-LSTM	0.00144	0.00197	0.07359	94.814
	KOA-LSTM	0.01647	0.02565	0.16931	92.221
PV Site 4	KOA-C-Bi-LSTM	0.0017	0.0045	0.00364	98.852
	KOA-C-LSTM	0.00814	0.00905	0.07324	97.063
	KOA-LSTM	0.01270	0.03134	0.15760	93.712

Table 6
Comparison of Statistical Analysis of Forecasting Results with Transductive Transfer Learning.

Dataset	Tech	RMSE	NMSE	MAE	R-Square
PV Site 5	TKCBL	0.0027	0.0014	0.00401	99.315
	TKCL	0.00760	0.00605	0.06769	96.011
	TKL	0.03509	0.01701	0.41683	93.103
PV Site 6	TKCBL	0.0059	0.0064	0.00196	99.515
	TKCL	0.00774	0.00409	0.07397	97.567
	TKL	0.01123	0.02985	0.51739	94.289
PV Site 7	TKCBL	0.0012	0.0050	0.00219	99.472
	TKCL	0.00208	0.00549	0.08225	96.889
	TKL	0.08467	0.07364	0.56102	93.467
PV Site 8	TKCBL	0.0013	0.0052	0.00107	99.657
	TKCL	0.00315	0.00418	0.01183	96.628
	TKL	0.03537	0.04462	0.20345	94.791

Table 7
Comparative Analysis with state-of-the-art techniques

Ref	Year	Technique	Error
[43]	2022	HBO-LSTM	R2 = 0.965 RMSE = 3.9%
[44]	2022	ACO-NN	R2 = 0.815 RMSE = 3.4%
[45]	2022	VMD-CNN-BiGRU	RMSE = 2.8%
[46]	2023	DSN	RMSE = 2.9%
[48]	2023	LSTM-TCN	RMSE = 1.153%
Our Tech.	2023	TKCBL	R2 = 0.9948, RMSE = 0.27%

Table 8
Comparison of models based on GCT average performance.

Model	F-value	P value	T statistic	Error Standard (⩽)
TKCBL	0.0879	0.0098	0.811	0.09
TKCL	0.1326	0.0421	1.201	0.14
TKL	0.1902	0.0924	1.823	0.20

H_0 : laggedvaluesof$X$donotGranger − cause$Y$

H_a : laggedvaluesof$X$doGranger − cause$Y$

In the Granger causality test, X represents the predictor variable and Y represents the target variable. The goal is to determine if past values of X can be used to forecast future values of Y.

Let's consider two time series, Y and X, with observations Y_t and X_t at time t.

The Granger causality model can be formulated as:

$$Y_t = \alpha_Y + \sum_i \beta_{Yi}\left(Y_{t-i}\right) + \sum_i \gamma_{Yi}\left(X_{t-i}\right) + \varepsilon_{Yt} \tag{57}$$

$$X_t = \alpha_X + \sum_i \beta_{Xi}(X_{t-i}) + \sum_i \gamma_{Xi}(Y_{t-i}) + \varepsilon_{Xt} \quad (58)$$

where α_Y and α_X are intercept terms, β_{Yi} and β_{Xi} are coefficients for the past values of Y and X, and γ_{Yi} and γ_{Xi} are coefficients for the past values of X and Y, respectively. ε_{Yt} and ε_{Xt} are error terms.

To test for Granger causality, we first estimate the above model using suitable estimation techniques, such as ordinary least squares (OLS). Then the F-statistic for the anthropic principle is calculated so that the coefficients γ_{Yi} are all equal to 0.

The F-statistic can be calculated as:

$$F = \frac{(RSS_R - RSS_U)/m}{RSS_U/(n-2m-1)} \quad (59)$$

where RSS-R is the residual sum of squares for the restricted model (i.e., where $\gamma_{Yi} = 0$ for all i), RSS-U is the residual sum of squares for the unrestricted model, m is the number of restrictions (like the number of γ_{Yi} coefficients), and n is the sample size.

The null hypothesis of no Granger causality is rejected if the F-statistic exceeds a critical value defined by the desired significance level and the degrees of freedom of the models. This indicates that past values of X have statistically significant predictive ability for future values of Y, or that X Granger causes Y. The results are summarized in Table 8.

5.4. Discussion

The results of our study demonstrate that the TKCBL model exhibits exceptional performance in accurately predicting PV generation across all eight PV sites. The consistently low values of RMSE, NMSE, and MAE indicate the model's ability to closely approximate the actual PV generation values, minimizing prediction errors. The superiority of the TKCBL model is evident as it outperforms the other two models, TKCL and TKL, by a substantial margin. Moreover, the high R-Square values achieved by the TKCBL model across all PV sites indicate a strong correlation between predicted and actual PV generation values, further emphasizing its effectiveness in capturing underlying patterns and trends in the data. These findings are particularly promising as accurate PV generation forecasting is essential for integrating large solar power plants with conventional grids. By achieving near-perfect explanatory power and significantly lower prediction errors compared to existing state-of-the-art techniques, the KOA-C-Bi-LSTM model represents a significant advancement in PV site data prediction and holds great promise for practical applications in the field.

The comparative analysis with existing techniques further solidifies the superiority of the proposed TKCBL model. Our model substantially outperforms other published techniques in terms of both R-Square values and RMSE, demonstrating its clear advantage in accurately predicting PV site output power. The TKCBL technique's ability to explain nearly all the variance in the target variable while maintaining an impressively low average prediction error sets it apart as a robust and reliable forecasting tool. The comparison highlights the substantial progress made in PV generation prediction, and the TKCBL model emerges as a standout performer. As PV power forecasting plays a crucial role in the integration of renewable energy sources into the grid, the exceptional accuracy and precision offered by our model can significantly enhance the efficiency and stability of the power system. The demonstrated advancements in forecasting accuracy, combined with the model's ease of implementation and scalability, position the TKCBL technique as a promising choice for future research and practical applications in the field of renewable energy integration. Moreover, the insights gained from this study may pave the way for further improvements in forecasting methodologies, enabling a more sustainable and reliable energy future.

6. Conclusion

This paper presents an innovative and comprehensive approach to enhancing the accuracy of power output forecasting in Photovoltaic (PV) power plants, particularly in dynamic environmental conditions. The Hybrid Deep Learning Model (DLM) developed in this research, which combines Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Bidirectional LSTM (Bi-LSTM), demonstrates the capability to capture both spatial and temporal dependencies within weather data, which are crucial for precise predictions. A significant contribution of this work is the introduction of the Kepler Optimization Algorithm (KOA) for hyperparameter tuning. Inspired by Kepler's laws of planetary motion, KOA optimizes the DLM's performance by identifying the most appropriate hyperparameter configurations, thus improving prediction accuracy. Additionally, the integration of Transductive Transfer Learning (TTL) into the deep learning models offers enhanced efficiency by leveraging prior knowledge to refine forecasting capabilities while minimizing resource consumption. The utilization of diverse datasets encompassing environmental parameters and PV plant-generated power across various sites enhances the robustness of the study. The proposed TKCBL achieves up to 99.65% R2 and 99.48% average R2 of all datasets. The average RMSE achieved by TKCBL in all datasets is 0.27% which is very low as compared to competing techniques. This outcome underscores the hybrid approach's potential to provide remarkably accurate and resilient power output predictions under diverse weather conditions.

The synergy of the hybrid DLM, KOA hyperparameter tuning, and TTL techniques offers a holistic solution to elevate the management of PV power plants. The ability to analyze spatial and temporal data effectively, coupled with efficient hyperparameter optimization and knowledge transferability, positions this approach as a valuable tool for ensuring efficient and dependable power output predictions within dynamic environmental contexts. These research findings contribute not only to the renewable energy field but also hold substantial promise for practical applications in PV power plant management. By leveraging these insights, stakeholders can make informed decisions and optimize the efficiency of PV power generation, further advancing the transition to sustainable and clean energy sources.

CRediT authorship contribution statement

Umer Amir Khan: Conceptualization, Methodology, Project administration, Validation, Data curation. **Noman Mujeeb Khan:** Conceptualization, Formal analysis, Investigation, Visualization, Data curation. **Muhammad Hamza Zafar:** Validation, Formal analysis, Supervision, Funding acquisition, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The link is shared for the data.

References

[1] Jacobson MZ. Review of solutions to global warming, air pollution, and energy security. Energy Environ. Sci. 2009;2:148–73.
[2] Cruz MR, Fitiwi DZ, Santos SF, Mariano SJ, Catalao JP. Multi-flexibility option integration to cope with large-scale integration of renewables. IEEE Trans. Sustainable Energy 2018;11:48–60.
[3] Hamada H, Kusayanagi Y, Tatematsu M, Watanabe M, Kikusato H. Challenges for a reduced inertia power system due to the large-scale integration of renewable energy. Global Energy Interconnection 2022;5.

[4] Reddy SS, Bijwe P, Abhyankar AR. Real-time economic dispatch considering renewable power generation variability and uncertainty over scheduling period. IEEE Syst. J. 2014;9:1440–51.
[5] Verzijlbergh R, De Vries L, Dijkema G, Herder P. Institutional challenges caused by the integration of renewable energy sources in the european electricity sector. Renew. Sustain. Energy Rev. 2017;75:660–7.
[6] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Solar Energy 2016;136:78–111.
[7] Hanifi S, Liu X, Lin Z, Lotfian S. A critical review of wind power forecasting methods–past, present and future. Energies 2020;13:3764.
[8] Ahmed R, Sreeram V, Mishra Y, Arif M. A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization. Renew. Sustain. Energy Rev. 2020;124:109792.
[9] Barbieri F, Rajakaruna S, Ghosh A. Very short-term photovoltaic power forecasting with cloud modeling: A review. Renew. Sustain. Energy Rev. 2017;75:242–63.
[10] Sahu RK, Shaw B, Nayak JR, et al. Short/medium term solar power forecasting of chhattisgarh state of india using modified tlbo optimized elm, Engineering Science and Technology, an. Int. J. 2021;24:1180–200.
[11] Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. Appl. Energy 2021;283:116239.
[12] Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. IEEE Trans. Power Syst. 2007;22:258–65.
[13] Mayer MJ, Yang D. Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting. Renew. Sustain. Energy Rev. 2023;175:113171.
[14] Barman M, Choudhury NBD. A similarity based hybrid gwo-svm method of power system load forecasting for regional special event days in anomalous load situations in assam, india. Sustainable Cities Soc. 2020;61:102311.
[15] Zhao Y, Ye L, Wang Z, Wu L, Zhai B, Lan H, Yang S. Spatio-temporal markov chain model for very-short-term wind power forecasting. J. Eng. 2019;2019:5018–22.
[16] Sivaneasan B, Yu C, Goh K. Solar forecasting using ann with fuzzy logic pre-processing. Energy Procedia 2017;143:727–32.
[17] Zhao Y, Ye L, Pinson P, Tang Y, Lu P. Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. IEEE Trans. Power Syst. 2018;33:5029–40.
[18] Xu X, Lu Y, Vogel-Heuser B, Wang L. Industry 4.0 and industry 5.0–inception, conception and perception. J. Manuf. Syst. 2021;61:530–5.
[19] Parmentola A, Tutore I, Costagliola Di Fiore M. Environmental side of fourth industrial revolution: The positive and negative effects of i4 technologies. In: Handbook of Smart Materials, Technologies, and Devices: Applications of Industry 4.0. Springer; 2021. p. 1–31.
[20] Leng J, Sha W, Wang B, Zheng P, Zhuang C, Liu Q, Wuest T, Mourtzis D, Wang L. Industry 5.0: Prospect and retrospect. J. Manuf. Syst. 2022;65:279–95.
[21] Huang S, Wang B, Li X, Zheng P, Mourtzis D, Wang L. Industry 5.0 and society 5.0–comparison, complementation and co-evolution. J. Manuf. Syst. 2022;64:424–8.
[22] Fang X, Misra S, Xue G, Yang D. Smart grid–the new and improved power grid: A survey. IEEE Commun. Surveys Tutorials 2011;14:944–80.
[23] Tuballa ML, Abundo ML. A review of the development of smart grid technologies. Renew. Sustain. Energy Rev. 2016;59:710–25.
[24] Lin KY. Dynamic pricing with real-time demand learning. Eur. J. Oper. Res. 2006;174:522–38.
[25] Hu Z, Kim J-H, Wang J, Byrne J. Review of dynamic pricing programs in the us and europe: Status quo and policy recommendations. Renew. Sustain. Energy Rev. 2015;42:743–51.
[26] Das L, Munikoti S, Natarajan B, Srinivasan B. Measuring smart grid resilience: Methods, challenges and opportunities. Renew. Sustain. Energy Rev. 2020;130:109918.
[27] Albasrawi MN, Jarus N, Joshi KA, Sarvestani SS. Analysis of reliability and resilience for smart grids. In: 2014 IEEE 38th Annual Computer Software and Applications Conference. IEEE; 2014. p. 529–34.
[28] Ahsan L, Baig MJA, Iqbal MT. Low-cost, open-source, emoncms-based scada system for a large grid-connected pv system. Sensors 2022;22:6733.
[29] Kermani M, Adelmanesh B, Shirdare E, Sima CA, Carnì DL, Martirano L. Intelligent energy management based on scada system in a real microgrid for smart building applications. Renewable Energy 2021;171:1115–27.
[30] Chen Y, Xu J. Solar and wind power data from the chinese state grid renewable energy generation forecasting competition. Scientific Data 2022;9:577.
[31] Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O. A survey on missing data in machine learning. J. Big Data 2021;8:1–37.
[32] Jin J, Li M, Jin L. Data normalization to accelerate training for linear neural net to predict tropical cyclone tracks. Math. Problems Eng. 2015;2015.
[33] Cohen I, Huang Y, Chen J, Benesty J, Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. Noise Reduction Speech Process. 2009:1–4.
[34] Urbanek T, Prokopova Z, Silhavy R, Vesela V. Prediction accuracy measurements as a fitness function for software effort estimation. SpringerPlus 2015;4:1–17.
[35] Abdel-Basset M, Mohamed R, Azeem SAA, Jameel M, Abouhawwash M. Kepler optimization algorithm: A new metaheuristic algorithm inspired by kepler's laws of planetary motion. Knowl.-Based Syst. 2023;268:110454.
[36] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. J. Big Data 2021;8:1–74.
[37] Liu Z, Wang H, Liu J, Qin Y, Peng D. Multitask learning based on lightweight 1dcnn for fault diagnosis of wheelset bearings. IEEE Trans. Instrum. Meas. 2020;70:1–11.
[38] Shahid F, Zameer A, Muneeb M. Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. Chaos, Solitons & Fractals 2020;140:110212.
[39] Nguyen THT, Phan QB. Hourly day ahead wind speed forecasting based on a hybrid model of eemd, cnn-bi-lstm embedded with ga optimization. Energy Reports 2022;8:53–60.
[40] Pinto G, Messina R, Li H, Hong T, Piscitelli MS, Capozzoli A. Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics. Energy Build. 2022;276:112530.
[41] M. Ren, E. Triantafillou, S. Ravi, J. Snell, K. Swersky, J.B. Tenenbaum, H. Larochelle, R.S. Zemel, Meta-learning for semi-supervised few-shot classification, arXiv preprint arXiv:1803.00676 (2018).
[42] Kushibar K, Salem M, Valverde S, Rovira À, Salvi J, Oliver A, Lladó X. Transductive transfer learning for domain adaptation in brain magnetic resonance image segmentation. Front. Neurosci. 2021;15:608808.
[43] Ewees AA, Al-qaness MA, Abualigah L, Abd Elaziz M. Hbo-lstm: Optimized long short term memory with heap-based optimizer for wind power forecasting. Energy Convers. Manage. 2022;268:116022.
[44] Netsanet S, Zheng D, Zhang W, Teshager G. Short-term pv power forecasting using variational mode decomposition integrated with ant colony optimization and neural network. Energy Reports 2022;8.
[45] Zhang C, Peng T, Nazir MS. A novel integrated photovoltaic power forecasting model based on variational mode decomposition and cnn-bigru considering meteorological variables. Electric Power Systems Research 2022;213:108796.
[46] Khan ZA, Hussain T, Baik SW. Dual stream network with attention mechanism for photovoltaic power forecasting. Appl. Energy 2023;338:120916.
[47] Rosoł M, Młyńczak M, Cybulski G. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. Comput. Methods Programs Biomed. 2022;216:106669.
[48] Limouni T, Yaagoubi R, Bouziane K, Guissi K, Baali EH. Accurate one step and multistep forecasting of very short-term pv power using lstm-tcn model. Renewable Energy 2023;205:1010–24.