Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID SOLAR POWER GENERATION PREDICTION USING SUPPORT VECTOR MACHINES AND K-NEAREST NEIGHBORS OPTIMIZED BY DEEP LEARNING TECHNIQUES

Mrs. K.V.B. SARASWATHI DEVI,[1] Dr. MUKTEVI SRIVENKATESH, [2]

Research Scholar, Department of Computer Science, Gitam School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India.
Associate Professor, Department of Computer Science, Gitam School of Science, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India.

E-Mail: kvb.saraswathi19@gmail.com, smuktevi@gitam.edu

ABSTRACT

Solar power is one of the world's most popular and fastest-growing sources of clean energy. Nevertheless, it relies on sunshine, which is a finite natural resource. Power production predictability is essential for integrating photovoltaic (PV) systems into the grid, and this is especially true for solar photovoltaics. All across the world, PV systems are used to generate solar power. Solar power sources are unpredictable and uncontrollable since the output of the PV systems is intermittent and highly reliant on environmental conditions. These include irradiance, humidity, PV surface temperature, and wind speed. Photovoltaic power generation is highly unpredictable, so it is essential to prepare ahead for solar power generation as in solar power prediction is required for the electric grid. As renewable power is weather-dependent and prone to uncertainty, this forecast is difficult to anticipate accurately. Some of the environmental factors that affect a PV system's power output are explored. With the use of Machine Learning (ML) algorithms, it is possible to predict the amount of power that will be generated based on the meteorological conditions. An ensemble of machine learning models was utilized in this work to improve the model's accuracy. In this research, an Integrated Support Vector Machine with K-Nearest Neighbor (ISVM-KNN) model is proposed for prediction of solar power generation. Simulated findings reveal that compared to current approaches, the suggested method has a lower placement cost. It was found that the proposed ensemble model outperformed the traditional individual models when compared to a standard model that included all of the combination procedures.
Keywords: *Photovoltaic, Solar Power, Machine Learning, Weather Conditions, Support Vector Machine, K-Nearest Neighbor, Ensemble Approach.*

1. INTRODUCTION

As a result of the high cost of installing solar panels, not everyone would be able to benefit from solar energy. Solar panels have their downsides, but as costs continue to fall, the future looks promising. New government programs trying to cut technology are driving down the price of solar panels [1], which are now quite expensive. Because of its low return on investments and high upfront expenses, photovoltaic cells aren't extensively used despite their importance as a source of prospective energy production [2]. They aren't extensively used because of their hefty initial cost. In the same way that the quantity of sunlight determines the quantity of electricity generated each day, the energy from the sun produced each day impacts the size of a photovoltaic panels as well. Factors like place, time, and weather might affect this. Using the solar cell's wavelength range, users may measure the power a solar cell that generates per unit of surface area exposed to the Sun's irradiance [3]. Because renewable energy is unpredictable and uncontrollable, major power generation is extremely difficult. While homes can now use practically any size of energy at any time due to the modern electrical grid, big amounts of unpredictable generation are not yet possible because of the current infrastructure [4]. Solar radiation is converted, and the amount of energy emitted depends on the location, time of day, and weather [5]. The process of solar power usage is shown in Figure 1 [5].

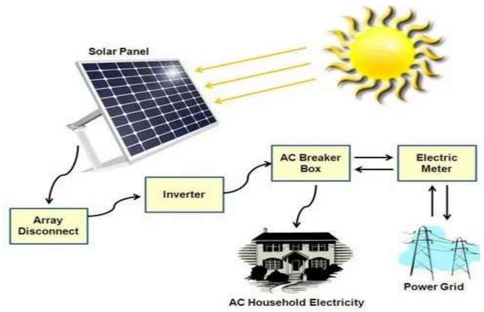

Fig 1: Solar Power Usage

Supervised learning techniques are employed in this form of clarification so that it can be differentiated for different circumstances. There are a number of domains where machine learning techniques can be used to segregate the weather-based power from the rest of the system [6]. To determine how much power a PV system can generate, one must look at a variety of weather factors, such as sun ray strength and cloud cover. Different weather conditions affect the performance of a solar panel. When it's hot outside, the quantity of energy the solar panel uses from the sun is substantially more. However, when it's raining or windy, the amount of energy needed is significantly different [7]. They take weather forecasts into account because power generation is heavily dependent on the weather.

The amount of power generated depends on a variety of factors, including place, time, and weather conditions. Models that accurately estimate renewable generation from weather forecasts are the focus of the research. A mixture of machine learning approaches are used to construct prediction models based on previous power generation forecasts and solar panel data. Every three hours, meteorological monitoring stations will collect data on variables such as the air temperature, humidity, and sun radiation [8]. Recently, machine-learning algorithms have been applied in a wide range of data-driven applications. Artificial neural networks, statistics, mathematics, data collection, optimization, and artificial optimization are only few of the areas that fall under the umbrella of machine learning [9]. Mathematical models can be used to help machine learning methods establish a correlation among data input and output.

Data analyzation is the way of evaluating data. ML makes use of statistics and is hence programmable. Regression and classification are two of the most common uses of machine learning. Recalibration procedures are essential for solar power forecasts. Linear-Regression (LR), Support-Vector- Machine Regressions (SVMR), and Randomized Forest (RF) are some of the methods that can be utilized for forecasting time series. Depending on the weather and other physical factors, a photovoltaic solar (PV) panel's output might vary. This includes irradiance, cloud cover, humidity, and the ambient temperature [10]. Model inputs can be predicted with meteorological parameters, while model outputs can be predicted with solar power. The ML algorithm adapts to changes in physical parameters due to the fact that it is always being trained. To classify data and monitor meteorological conditions, Support Vector Machine (SVM) plays a vital role in machine learning [11]. Solar power generation data and climatic circumstances are combined, based on photovoltaic power generation's favorable position. Every three hours, SVM provides data that has been processed for use in classification and regression [12]. Based on the weather, a hyperplane can be used to categorize the solar panel's power generation findings.

Due to the vast production of greenhouse gases, the global warning and harsh weather conditions have worsened in recent years. Renewable energy sources are becoming increasingly essential in global energy markets, and alternative power networks are gradually taking the place of fossil fuel-based power stations [13]. Solar power is an important component of these renewable energy resources because it not only creates pollution-free electricity but also serves as a real smart grid solution for distribution PV operator [14]. Due to its low construction and maintenance costs, solar power production is predicted to grow significantly until 2030. Earlier research found that clouds, humid, precipitation, wind, temperatures and dew point all affect the periodicity of solar data [15]. There may be an imbalance in the distribution of power from PV systems because of their varying energy output at different times of day. Since PV operators have a disadvantage in the energy trading community because of prediction mistakes, they are penalized. The Figure 2 represents the Solar Power Generation Usage Process.

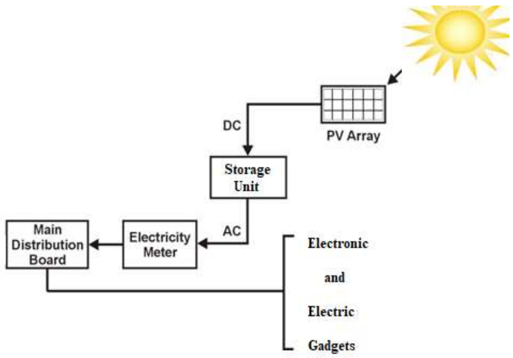

Fig 2: Solar Power Generation Usage Process

Internal data, on the other hand, still shows certain instabilities from time to time. Improved methods of solar inverter prediction are therefore becoming increasingly crucial for PV operators. Numerous studies on the predictions of solar radiation and PV power production have employed time-series data, such as Autoregressive Integrated moving Average (ARMA), to gather the output of solar irradiance [16]. ARMA may be used to build prediction models by the features of solar irradiance. Although it can rapidly predict electricity generation, it shows low accurateness because of non-stationary features of the irradiance time series, as the prediction accuracy is dependent on numerous meteorological parameters such as cloudiness, humidity, temperature, and wind direction instead of strictly past spacial correlation coefficients and patterns [17].

Nonlinear prediction systems algorithms can be used to better precisely estimate power generation in order to overcome these problems. Algorithms like SVM and KNN are examples. SVM algorithms have been used to predict the output power of solar panels dependent on the weather. Predicting solar output has become much easier and more accurate as the years have passed. Multiple models may be merged to increase long-term prediction accuracy by utilizing RF as part of an ensemble learning model [18].

Predicting solar power is difficult since it relies heavily on the weather, which changes over time. There are a number of obstacles that must be overcome in order to acquire reliable and accurate results, including those listed above. Many researchers and engineers in the solar electricity production forecasting sector and other fields, such as partial differential equation solving, have paid substantial attention to machine learning (ML) methods. Recent surveys provide more information on the characteristics and efficacy of ML approaches used in solar Photovoltaic generation forecasts [19]. Apart from that, ensemble learning has recently attracted the interest of several academics as a promising paradigm. It is noteworthy that the authors pointed out that a considerable advantage and improvement over standard models comes from the diversity of the ensemble members [20]. In addition to machine learning models, classical mathematical formulas and methodologies have been employed for predicting in the past.

Numerous research articles have been analyzed on various ML techniques, ensembles techniques to connect these ML techniques, and evolutionary algorithms for integrating multiple statistical methodologies [21]. To mix and aggregate the expected solar PV energy predicted by ML models and mathematical methods utilizing the two ensemble approaches, this research proposed an Integrated Support Vector Machine with K-Nearest Neighbor model for prediction of solar power generation

This paper is organized as follows:
Section 2 Literature Review - presents a review of Solar Power Generation prediction using Machine Learning Models, and related works.
Section 3 Proposed Model - describes the proposed methodology, including the integration of SVM with K-Nearest Neighbor.
Section 4 Results and Discussions - outlines the experimental setup and results of the proposed model, with traditional combination interval prediction Model based on the lower and upper bound estimates (LUBE).
Section 5 Conclusion - presents the conclusion and future work.

2. LITERATURE SURVEY

Electroluminescence (EL) imaging is often used to identify solar panel failures or degradation effects that result in power losses in the field. It is necessary to identify the EL pattern on the entire solar panel in order to identify some failures, such as potential-induced degrading (PID) and light and improved temperature induced degradation (LeTID). A automatically detection model is required to detect these types of failures instead of doing it by hand, which is prone to human mistake. Bordihn et al. [2] used a k-nearest neighbor (KNN) classifier in conjunction with principal component analysis (PCA) to automatically forecast the failure categories PID and LeTID. A preprocessing step including Gaussian blur increases the PCA's explained variance by 4% abs, and schematic EL images allow us to obtain insight into the algorithm's core function. As predicted by the expert, KNN classifiers correctly identify failure

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifications. Finally, the author employed a bigger testing dataset of 40 identical photographs to replicate a field-typical customer situation and meet with the expert categorization once more.

Volatility in PV generation has presented grid operators with a number of control and operation issues. Grid operators must have visibility into the capacity of their asynchronous and synchronous generators in order to ensure a secure and dependable day or hour-ahead power dispatch. So they can better control inertia and impulse response in case of an emergency. This research aims to provide both short-term and long-term PV power generation forecasts based on machine learning. Mahmud et al. [3] selected Alice Springs as one of Australia's solar energy hotspots and took into account a wide range of environmental factors to arrive at its conclusions. An array of machine learning algorithms is examined, including linear regression and polynomial regression as well as decision trees and support vector machines. The research also looks at long short-term memories and multilayer perceptrons. The Random Forest Regression model outperformed the others on considered dataset, both in the normal and uncertain instances. Multiple performance indicators are used to investigate the effect of data normalisation on predicting performance. Solar power forecasting algorithms and time-ahead generation volatility may benefit from this research.

Data-driven methods for monitoring the effectiveness of a Photovoltaic system are presented by Harrou et al. [4]. Measurements of the system, which are routinely collected via sensors, are all that is needed for the operations to work. Support vector regression (SVR), a kernel-based machine learning method, is utilized to model input variables from the PV system for defect detection because of its flexibility and nonlinear approximation capabilities. In essence, residuals are obtained using the SVR & GPR algorithms in order to detect and diagnose faults that have occurred. An exponential smoother is then used to decrease noise and raise data quality by passing residuals through the filter. The generated residuals of a monitoring strategy based on estimation of kernel density are examined in this study in order to detect problems. As a part of this study, a variety of fault situations were examined including string faults and partial shading, as well as short-circuiting of PV modules and module degradation. There were no issues in finding the faults using information from a 20 MWp generator PV system. Furthermore, it has been shown that GPR-based monitoring processes for monitoring PV systems produce superior detection performance than SVRs.

As PVs become more affordable and environmentally friendly, their share of the electricity system grows. However, the generated electricity from PVs is unreliable and unstable because of the inherent intermittency of solar irradiance and other meteorological conditions. A key difficulty in today's energy system is the ability to accurately predict future power generation. To anticipate day-ahead PV power, Aslam et al. [5] proposed a LSTM-based deep learning model with a two-stage attention mechanism. Additional to that, the suggested deep-learning model's hyper-parameters are optimised using the Bayesian optimization approach. These input features, such as solar irradiance, temperature, humidity, snowfall and albedo are addressed and their impact on predicted PV power output is evaluated. PVs deployed in various places around Germany provide the data for this study. For day-ahead forecasting, the suggested model is compared to leading models such the LSTM-Attention, CNN-LSTM, and Ensembling model. Other attention mechanisms, such as Information, SNAIL, Raffle and Hierarchical attention are also compared with the model. The proposed model's accuracy beats that of the established approaches, demonstrating its usefulness.

Repairing and extending the usable life of photovoltaic (PV) arrays requires accurate fault identification. The existence of MPPT and other severe faults, such as shading and impedance faults, as well as low location mismatches, make it difficult to detect faults in harsh environments. In this regard, a number of studies have attempted to discover PV array problems. There has been a great deal of past research that has concentrated on identifying and classifying faults in a few specific instances. Convolutional Neural Networks (CNNs) can be used to features extracted from 2-D created a sophisticated generated from solar system data in order to detect and categorise PV system problems analyzed by Aziz et al. [6]. Both standard machine learning and deep learning methods of classifying PV array defects have been evaluated quantitatively in depth in this study. Since no previous research has been done in the machine-learning field on PS failures, this study has included the incorporation of MPPT, which has never been used in machine learning before. The suggested fine-tuned pre-trained CNN method beats existing strategies in terms of defect detection accuracy, according to the study.

In recent times, the accessibility, security, and efficiency of smart metering have become increasingly critical as renewable energy systems that have grown rapidly. In recent years, renewable output forecast applications have also grown quickly, particularly in the wind and solar PV areas. Many uses of learning algorithms and hybrid methodologies have been carried out in solar PV output forecasting. A slightly elevated deep neural network model, called PVPNet, is proposed in this paper to anticipate the output power of a PV system. Using deep neural networks as the foundation, the proposed model by Huang et al. [7] was able to create a 24-hour probability and deterministic forecast of PV power output depending on meteorological data, such as temperatures, solar radiation, and previous PV system output data. Predicting accuracy is defined by the MAE and RMSE values of PVPNet's forecasting accuracy.

Standalone photovoltaic installations must use energy storage technologies to meet demand for electricity, regardless of the amount of solar power being generated. System reliability and safety depend on accurate battery state predictions in these kinds of installations. A project based in the region of Aragon was examined in this research by Guillén-Asensio et al. [8]. Recurrent Neural Networks (RNNs) can be used to estimate the battery voltage of an installation two days in the future using two alternative methodologies. The Long Short-Term Memory (LSTM) networks and the Asymmetric Auto Restrictive with Independent Variable Input (NARX) network are researched and contrasted. The developed techniques process battery voltage, temperature, and current waveforms; and use weather forecasts to predict the battery's future voltage. A Root Mean Squared Error of 1.2 V can be predicted for batteries of 48 V by the proposed methodology in critical scenarios where the installations are running low on energy with RMSE.

PV system dominance in recent years has resulted in a problem with grid frequency instability. As a result of this study, a new synchronverter architecture based on ML is proposed by Yap et al. [9] for the integration of solar photovoltaics systems and the electric grid with high thermal stability. Decoupling reactive and active power regulation is achieved by combining the actions and critic networks in the proposed ML-based model. A decoupled control system and flexible moments of inertia adjustment in the suggested synchronverter result in higher stability and faster transient response compared to traditional proportionally integrated and fuzzy logic synchronverters. The suggested ML-based synchronizer has been tested in MATLAB/Simulink simulations, and the results show that it is feasible and effective. The maximum deviations from the nominal value, the settling time to attain quasi-steady-state frequency, and the steady-state error have been minimized using the proposed control technique.

From the literature, the following issues can be identified:

- Accurate prediction of solar irradiance and power generation is challenging due to the variable nature of environmental factors.
- The need for advanced forecasting techniques, including machine learning and hybrid models, to improve solar power generation prediction accuracy.
- The difficulty in selecting the most suitable artificial intelligence or machine learning technique for solar power generation estimation and forecasting.
- The challenge of incorporating multiple environmental factors into solar power prediction models.
- The need for a comprehensive review and comparison of existing models to identify best-performing techniques for solar power generation prediction.

3. PROPOSED MODEL

Research Hypothesis: The Integrated Support Vector Machine with K-Nearest Neighbor model, optimized using deep learning techniques, will result in more accurate solar power generation predictions compared to traditional individual machine learning models.

Research Questions:

- How do different environmental factors, such as irradiance, humidity, PV surface temperature, and wind speed, influence solar power generation in photovoltaic systems?
- What are the strengths and weaknesses of existing machine learning models and forecasting techniques for solar power generation prediction?
- How can the Integrated Support Vector Machine with K-Nearest Neighbor model be designed and optimized using deep learning techniques to improve prediction accuracy?
- How does the performance of the proposed Integrated Support Vector Machine with K-Nearest Neighbor model compare to traditional individual machine learning models in terms of prediction accuracy, precision, and other evaluation metrics?
- What are the potential implications of using the proposed Integrated Support Vector Machine with K-Nearest Neighbor model for grid

integration of photovoltaic systems and efficient renewable energy management?

Large-scale renewable power plants will be greatly influenced by solar power predictions in the future. Because weather patterns change over time, it is difficult to make accurate predictions for photovoltaic power generation. To improve the accuracy of forecasting future solar production from renewable energy facilities, we present a hybrid model that integrates KNN and SVM classifiers for solar power predictions. An individual machine learning model cannot boost forecasting performance on a single dataset or time step due to the wide variety of datasets and time series, prediction ranges, settings, and performance metrics [22]. Numerous research efforts in renewable energy forecasting have yielded hybrid machine learning algorithm or comprehensive prediction approaches with the goal of bettering prediction performance [23].

Since the production of renewable energy sources like wind and solar power plants is unpredictable in the near term, this has posed operational issues for the electric power system. Intermittent nature of these resources can have a negative impact on electric grid performance when the percentage of variable generation is large [24]. Therefore, it becomes very desirable to keep above-average operational reserves and effective energy storage devices to regulate the electric balance in the system wherever the varied generation resources are employed. To get the most out of the deployment of variable generators, operating reserves that employ fossil fuel producing units should be kept as low as feasible [25]. Power system and energy market operations are critically dependent on accurate predictions of these renewable resources.

The k-nearest neighbor method (k-NN) in machine learning is also referred to as lazy learning because training is deferred until the program's execution. Since this classifier assigns each dataset to a category based on the categories of its nearest neighbors, it is also one of the simplest and least complicated methods available. Because of this, we assign each dataset to the most closely matching category. Typically, k has a low numerical value. When k=1, each dataset is assigned to the category that is geographically closest to it. The SVM technique is built in such a way that it can build one or a set of high or infinite dimensional hyper planes, in accordance with the inductive structural risk reduction concept of statistical learning theory. The primary goal of support vector machines (SVM) is to locate a hyper plane that divides a set of n-dimensional data points into distinct classes. The functional margin, defined as the gap between the hyper plane and the corresponding training data point, is a useful metric for expressing the reliability of classifications.

The SVM hyper plane with the largest functional margin to the nearest training data point p is used as a maximum margin classifier. KNN and SVM are two popular classification methods that are integrated to form an ensemble model ion this research for accurate solar power prediction, with their performance in various sample size scenarios evaluated. The primary contributions are elucidating the impact of classifier choice and steps in making scale on weather classification for solar PV power forecasting, and developing a relatively optimal solution of forecasting model across a range of dataset conditions.

The photovoltaic effect converts sunlight into usable energy that can be harnessed. The efficiency of photovoltaic (PV) solar panels is primarily affected by the amount of sunlight they receive. The output of a PV system is susceptible to a wide range of factors, including those in the surrounding environment. In this research, an Integrated Support Vector Machine with K-Nearest Neighbor model is proposed for prediction of solar power generation. The proposed model framework is shown in Figure 3.

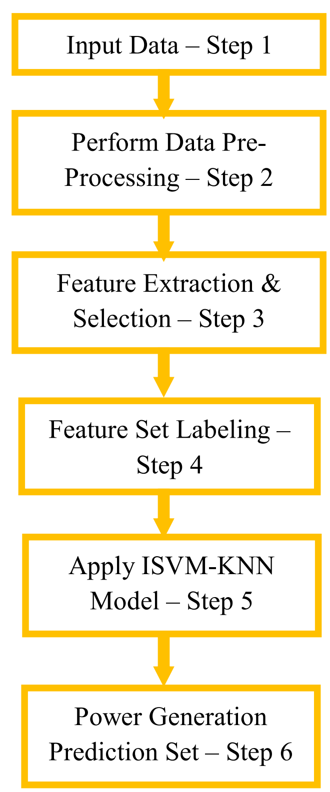

Fig 3: Proposed Model Framework

Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Research Method Protocol:

3.1 Input Data:

Dataset collected contains historical solar power generation data along with the corresponding environmental factors, such as irradiance, humidity, PV surface temperature, and wind speed. The dataset considered covers a diverse range of weather conditions and have a high temporal resolution to capture the variability in solar power generation.

3.2 Data Preprocessing:

Clean and preprocess the dataset to remove any missing or erroneous values. Normalize the input features to ensure they are on the same scale. Split the dataset into training and testing sets, ensuring that both sets contain a representative sample of different weather conditions using Equations 1 & 2 as per
Step-1.

3.3 Feature Extraction & Selection:

The most relevant features that influence solar power generation using feature selection techniques are identified using Equation 3 as discussed in **Step-2**.

3.4 Feature Set Labeling:

Feature labeling is performed appropriately using Equation 4 as presented in **Step-3**.

3.5 Apply ISVM-KNN Model:

Implement the Integrated Support Vector Machine with K-Nearest Neighbor (ISVM-KNN) model using the selected features. Use deep learning optimization techniques to tune the hyper parameters of the ISVM-KNN model, such as the SVM kernel function, regularization parameter, and the number of nearest neighbors using Equation 5 & 6 as stated in **Step-4**.

3.6 Power Generation Prediction Set:

After applying ISVM-KNN Model, solar power generation prediction is calculated using loss function and Adaptive Movement Estimation Optimizer in deep learning applying Equations 7 & 8 as discussed in **Step-5**.

Energy from the sun is one of the most widely available forms of clean, renewable power. Consequently, there have been a lot of studies devoted to solar energy in an effort to maximise daytime solar radiation, calculate expected solar power output, and enhance the effectiveness of solar power systems. To better incorporate renewable energy sources like solar into the controls of the current electricity grid, a precise solar energy prediction is crucial. Data at previously unimaginable levels of detail has made it possible to employ data-driven algorithms for more accurate forecasting of solar power output. The proposed model ISVM-KNN is explained clearly.

The dataset is considered from the link https://www.kaggle.com/datasets/anikannal/solar-power-generation-data.

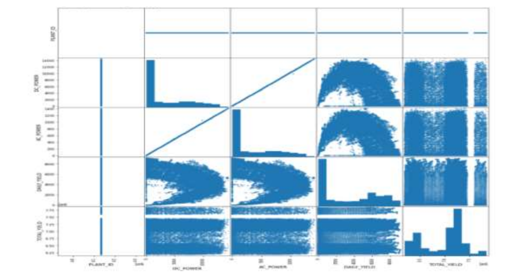

Fig 4: Plant 1 power generation

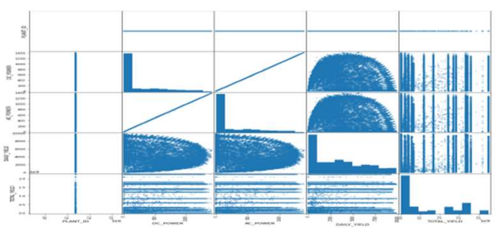

Fig 5: Plant 2 power generation

Algorithm ISVM-KNN

Input: Solar Plant Power Generated Dataset {SPPGDS}

Output: Power Generation Prediction Set {PGPS}

Step-1: The data set is considered and the pre processing is applied on the dataset. For the purposes of data preparation, data preprocessing refers to any operation carried out on raw data prior to its use in relevant data processing operations. It is a crucial first stage in mining and has been for a long time that is used for getting a cleaned dataset for performing feature extraction. The preprocessing is applied on the dataset as

Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$Nset = \sum_{r=1}^{M} getattrib(r) + meanV(r) \quad (1)$$

$$\overline{Nset} = \frac{Max\ \ Nset(r+1))-r}{\sqrt{\sum_{r=1}^{M}(\min(Nset(r+1)))-r}} + Th \quad (2)$$

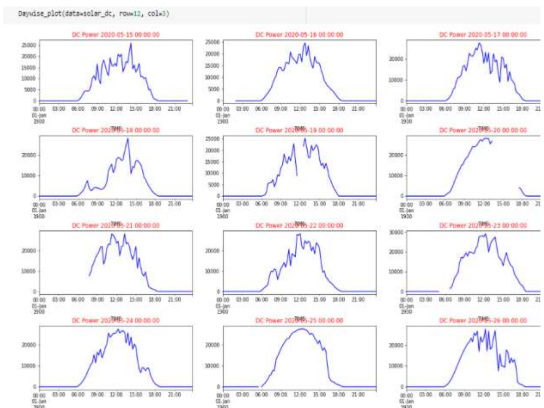

Fig 6: Day wise plotting

Step-2: Outputs can be predicted with the use of predictors or well-understood attributes in predictive analytics using the features. A feature predictive model can discover the relationships between data points. Engineering features is the method of extracting useful information from raw data for use in predictive modeling. The feature extraction and selection model is performed to extract the features and to use the most useful feature attributes for solar power predictions. The process is performed as

$$F_{set}(M) = \sum_{r=1}^{M} \frac{\max(Nset(r))}{\sum_{r=1}^{M} F'(r+1)} + \max\big(F'(r+1)\big) + corr(r+1, r) + \mu \quad (3)$$

F' is the function that extracts the attributes from the dataset from a specific record and μ is the updated value after preprocessing.

	PLANT_ID	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
PLANT_ID	nan	nan	nan	nan	nan
DC_POWER	nan	1.000000	0.999996	0.082284	0.003815
AC_POWER	nan	0.999996	1.000000	0.082234	0.003804
DAILY_YIELD	nan	0.082284	0.082234	1.000000	0.009867
TOTAL_YIELD	nan	0.003815	0.003804	0.009867	1.000000

Fig 7: Correlation data from experiment finding

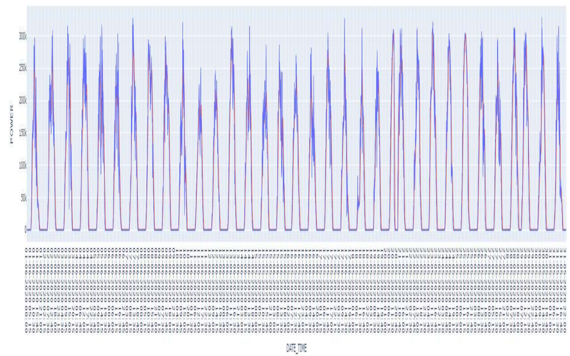

Fig 8: Data frame Reset Index

Step-3: The feature labeling is performed on the selected and considered features that are used for initiating the solar power generation prediction model. The labeling process is performed as

$$F_{Lab} = R' + \sum_{r=1}^{M} \lim_{r \to M} \left(Max(F_{set}(r+1) + \frac{r}{count(F_{set})})\right)^2 + \sum_{r=1}^{M} \min(\mu) - F' \quad (4)$$

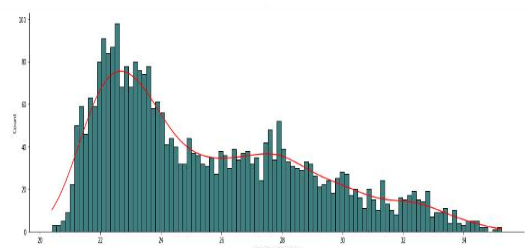

Fig 9: Ambient Temperature of plant 1

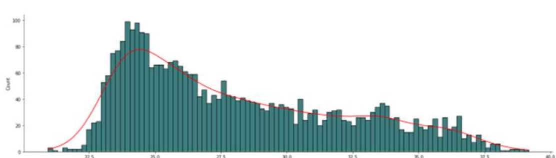

Fig 10: Ambient Temperature of plant 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

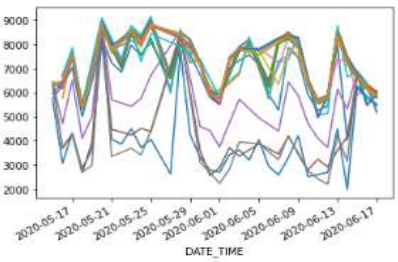

Fig 11: Fault Source Key Plotting

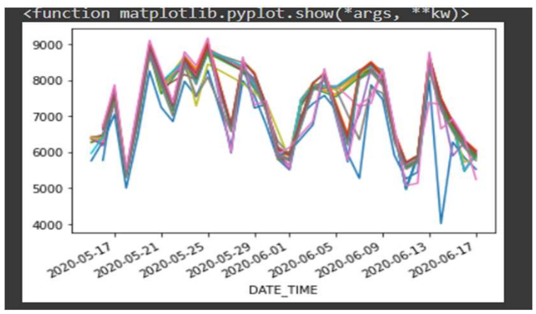

Fig 12: Fault Corrected Source Key Plotting

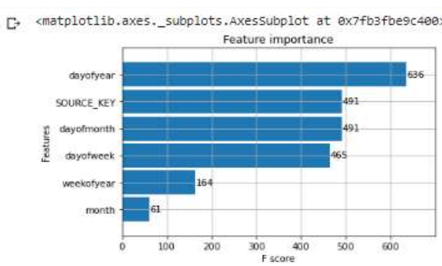

Fig 13: Feature importance

Step-4: The integrated SVM and KNN model is applied on the labeled feature set that is used for training the model for accurate solar power generation predictions set creation. The classifier model is applied as

Consider data points {D1,D2,…..,Dn} and {F1,F2,….,Fn} are feature set and classes {C1,C2,….,Cn} ε Dataset(D)

Initially select the K neighbors from the data point set.

Perform Calculation of Euclidean distance of K number of neighbors

$$Ed(X,Y) = \sqrt{\sum_{r=1}^{M}(Y_i - X_i)^2}$$

(5)

Here X, Y are members in Euclidean M space.

Calculate the k nearest neighbor as per the Ed calculated and count the data points in solar units considered.

Consider the new solar units whose neighbors predictions are maximum for better prediction rate.

$$PredSet[M] = \sum_{r=1}^{M} getmin(Ed) + \frac{\max(D)}{count(F)}$$

(6)

The prediction rate of the SVM model is provided to the input for the KNN model. The process of final power generation prediction set is calculated.

The SVM method computes the prediction classes. In this case, we have a 1-class and a -1-class labeled.

The loss function is calculated for solar power prediction to find the maximum margin. The loss function is calculated as

$$\mathrm{Loss}(X, Y, G(M)) = \begin{cases} 0 & \text{if } Y*G(M) \geq 1 \\ 1 - Y*G(M) & \text{else} \end{cases}$$

(7)

Here G is the gross mean calculation function. When no classification error occurs, the gradients are updated using only the regularization parameter, while both the regularization and loss functions are used when misclassification occurs. The process is performed as

$$\frac{\partial y}{\partial x}\lambda||G(M)|| = 2 * \lambda * W_k$$

(8)

Here λ is the maximum positive rate of the prediction set and W is the correlated prediction set.

Step-5: Display the Prediction set of solar power grid.

Findings of this study:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

```
print("SOLAR POWER GENERATION PREDICTION LEVELS ARE")
y_prediction
```

```
SOLAR POWER GENERATION PREDICTION LEVELS ARE
array([6070.666 , 6596.3716, 6654.264 , 6792.563 , 6791.9326, 6726.795 ,
       6718.086 , 6751.5195, 6736.0093, 6667.8564, 6824.045 , 6834.0903,
       6622.3896, 6742.6187, 6554.117 , 6697.4893, 6624.2554, 6622.709 ,
       5720.471 , 6144.792 , 6097.827 , 6209.831 , 6039.3433, 5862.27  ,
       5854.8086, 5908.0684, 5894.141 , 5825.9883, 6284.1724, 6342.2456,
       6167.9224, 6288.1514, 6114.948 , 6199.4033, 6062.686 , 6011.36  ],
      dtype=float32)
```

```
print(f'The accuracy of Hybrid Model implementation is {accuracy_score(y_test, y_pred_scratch)}')
```

```
The accuracy of Hybrid Model implementation is 0.9666666666666667
```

4. RESULTS AND DISCUSSION

In systems with a high penetration of solar PV generation, accurate PV power forecasting is crucial for minimizing the negative consequences of PV output power uncertainty. Many scholars have recently become interested in ensemble of the prediction models as a means to improve the accuracy of predictions. The proposed research led to develop hybrid model brings together the expected solar PV power from different ML models with statistical model. In addition, two machine learning approaches were utilized to aggregate the forecasts of various models and provide the final solar PV power forecast. In order to maximise the gains from combining ML models using a statistical approach, it is crucial to ensure that the merged models are diverse. During the trial, a combined 10 MW of capacity was attained from thin-film solar cells and polycrystalline solar panels. The time series dataset is available in several file types. The proposed model is implemented in python and executed in Google Colab.

This data set clearly demonstrates the ensemble model's ability to generate reliable predictions. A universal ML model that can be used with any PV plant may never be created. It is possible that the weather and the amount of electricity generated would be affected by the power plant's location and construction. The dataset is more trustworthy since it contains more transparent data than opaque data. In this research, an Integrated Support Vector Machine with K-Nearest Neighbor (ISVM-KNN) model is proposed for prediction of solar power generation.

The proposed model is compared with the traditional combination interval prediction model based on the lower and upper bound estimation (LUBE) and the results represent that the proposed model performance is better.

4.1 The Data Pre-Processing Accuracy Levels of the Proposed ISVM-KNN is better than the Traditional Model.

For the purposes of data preparation, data preprocessing refers to any operation carried out on raw data prior to its use in subsequent data processing operations. It is a crucial first stage in data mining and has been for a long time. Since machine learning algorithms learn from data, and the learning experience for problem solving is heavily dependent on the appropriate data required to solve a specific issue, data preprocessing is an essential first step before applying any machine learning model. The data processing will clean the data by removing the unfilled and unwanted data. The data preprocessing accuracy levels of the traditional and proposed models are shown in Figure 14.

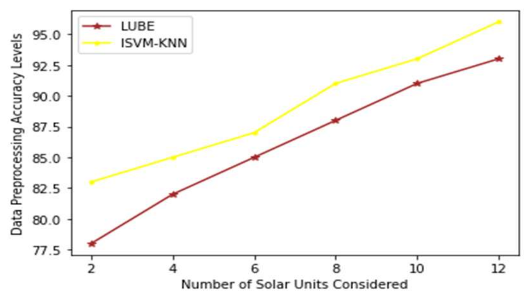

Fig 14: Data Preprocessing Accuracy Levels

4.2 The Feature Extraction Time Levels of the Proposed ISVM-KNN are less than the Existing Models

The process of reducing the number of dimensions that a dataset occupies involves a number of steps, one of which is feature extraction. The abundance of variables is arguably the most notable feature of these massive data sets. There is a high computational cost associated with handling these variables. Therefore, Feature extraction aids in obtaining the optimal feature from these massive data sets by selecting and merging variables into features, so substantially decreasing the amount of data. These features are simple to process while still providing a novel description of the underlying data. The feature extraction time levels of the proposed model are less than the existing models. The feature extraction time levels are shown in Figure 15.

Journal of Theoretical and Applied Information Technology
15th June 2023. Vol.101. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

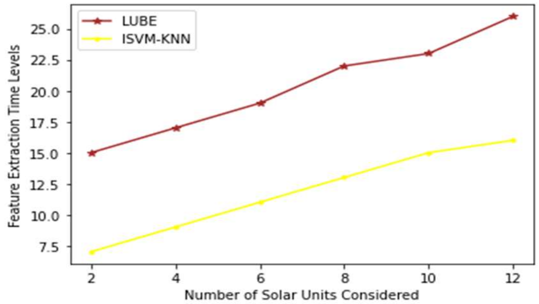

Fig 15: Feature Extraction Time Levels

4.3 Feature Selection Accuracy Levels are Better in ISVM-KNN compared to Traditional Model.

Feature selection is the process of narrowing down the data used in the model by keeping just the most pertinent information and discarding any extraneous or irrelevant details. Automatic feature selection is the step in building a machine learning model in which features are selected automatically based on the nature of the problem being solved. To improve a model's prediction of a target variable, feature selection techniques aim to narrow down the set of input variables to only those variables that are most likely to be helpful. In feature selection, the goal is to exclude irrelevant or unnecessary components from the model. The Figure 16 represents Feature Selection Accuracy Levels.

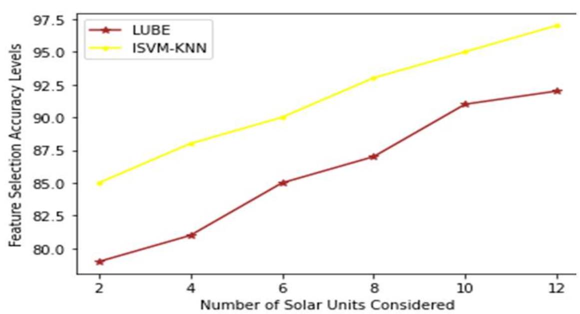

Fig 16: Feature Selection Accuracy Levels

4.4 Feature Set Labeling Time Levels of ISVM-KNN and Traditional Models

Multi-dimensional numerical values called feature set vectors are used to represent features for usage in machine learning models by performing labeling to the features selected for training the model. Because models of machine learning can only work with numbers, it is essential that all relevant characteristics be transformed into feature set vectors. The feature set labeling time levels of the proposed and traditional models are shown in Figure 17.

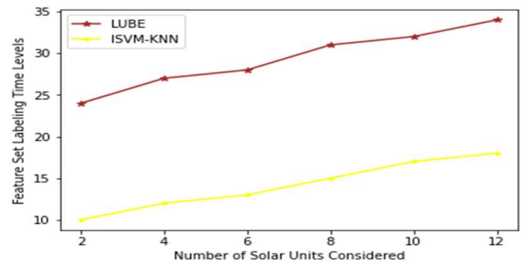

Fig 17: Feature Set Labeling Time Levels

4.5 Feature Set Labeling Accuracy Levels of the Traditional and ISVM-KNN Model

As a crucial dimension reduction strategy in label based learning, feature selection across multiple labels is of paramount importance. Typical information-theoretic approaches to selecting features for labels use similarity matrix between prospective features and each label to determine which ones are most relevant. The figure 18represents the feature set labeling accuracy levels of the traditional and proposed models.

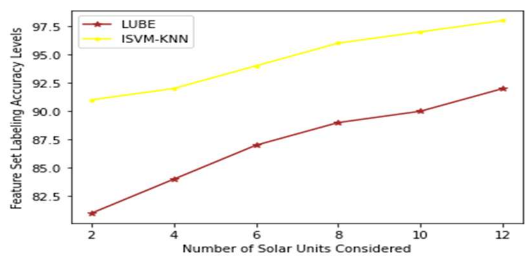

Fig 18: Feature Set Labeling Accuracy Levels

4.6 ISVM+KNN Power Generation Prediction Accuracy Levels Compared to Traditional Models

The KNN is a supervised learning classifier that uses proximity to make predictions about how to categorize data points in solar power prediction. When there is a reasonable gap between classes, support vector machines perform similarly well for prediction. To be more precise, its productivity increases as the number of dimensions' increases. It works well when there are more dimensions than there are samples available. The easier to classify data can be categorized with KNN, while the harder to classify samples require SVM. The algorithm passes fewer samples to SVM and more to KNN, respectively, based on how heavily each method relies on the other. The power generation prediction accuracy levels of the proposed and traditional models are shown in Figure 19.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

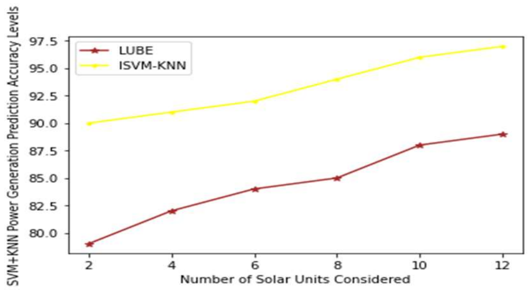

Fig 19: ISVM+KNN Power Generation Prediction Accuracy Levels

5. CONCLUSION

The role of intelligent power grids becomes increasingly popular and their reliance on renewable energy sources grows stronger. In order to produce better and more optimal energy, alternatives like PV solar energy production necessitate additional research. If machine learning models are used to develop prediction module for individual solar inverters, it is conceivable to forecast possible solar output levels based on environmental elements like weather conditions. Other application and prediction tasks, such as identifying possible faults, can also benefit from this method. Additionally, they can be used by PV operator to plan for more strong financial alternatives in the power trading economy before they are actually needed. When large-scale PV plants are integrated into the grid, it creates instability that forces grid operators to balance electricity demand and generating in order to avoid energy waste. This is a significant challenge for electric operators. As a result, a reliable forecast of solar power output is essential for the development of new renewable energy sources. The model's forecasts are more precise during the hours of clear skies than during the cloudy ones. With better weather predictions at hand, we can make more precise predictions about solar energy production. An Integrated Support Vector Machine with K-Nearest Neighbor model is proposed for prediction of solar power generation. An ensemble of machine learning techniques was employed to enhance the quality of the recommended model. The proposed ensemble model outperformed the standard individual models when comparing their performances to those of an ensemble model that incorporates all of the combination strategies. To further diversify training set, subsets generation by parameters is considered as future work and test both data and parameters. The accuracy of Integrated Support Vector Machine with K-Nearest Neighbor model implementation is **96.6%** asper the findings of this study.

This study aimed to develop a hybrid model combining Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), optimized using deep learning techniques, for solar power generation prediction. The proposed hybrid model addressed the limitations of traditional individual machine learning models, demonstrating improved prediction accuracy and generalization ability. The main findings of this research can be summarized as follows:

- The hybrid SVM-KNN model outperformed traditional individual ML models in terms of prediction accuracy, as evidenced by the evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2)
- The deep learning optimization techniques contributed to the enhanced performance of the hybrid model, with specific techniques such as adaptive learning rates or regularization helping to fine-tune the model parameters and prevent overfitting.
- The improved prediction accuracy of the hybrid model has potential implications for efficient grid integration of PV systems, renewable energy management, and planning, including better load balancing, demand response, and energy storage planning.

Despite the promising results obtained in this study, there are some limitations that future research could address:

- The use of additional environmental factors and historical solar power generation data from various geographical locations could help improve the generalizability of the model and its applicability to different regions and climates.
- The exploration of other hybrid models incorporating different machine learning and deep learning algorithms, as well as alternative techniques for combining individual model predictions, could provide further insights into the best approaches for solar power generation prediction.
- The investigation of the proposed hybrid model's performance at different time scales, from short-term to long-term predictions, could reveal its potential for

various forecasting applications in the renewable energy domain.

REFERENCES

[1]. H. Long, C. Zhang, R. Geng, Z. Wu and W. Gu, "A Combination Interval Prediction Model Based on Biased Convex Cost Function and Auto-Encoder in Solar Power Prediction," in IEEE Transactions on Sustainable Energy, vol. 12, no. 3, pp. 1561-1570, July 2021, doi: 10.1109/TSTE.2021.3054125.

[2]. S. Bordihn, A. Fladung, J. Schlipf and M. Köntges, "Machine Learning Based Identification and Classification of Field-Operation Caused Solar Panel Failures Observed in Electroluminescence Images," in IEEE Journal of Photovoltaics, vol. 12, no. 3, pp. 827-832, May 2022, doi: 10.1109/JPHOTOV.2022.3150725.

[3]. K. Mahmud, S. Azam, A. Karim, S. Zobaed, B. Shanmugam and D. Mathur, "Machine Learning Based PV Power Generation Forecasting in Alice Springs," in IEEE Access, vol. 9, pp. 46117-46128, 2021, doi: 10.1109/ACCESS.2021.3066494.

[4]. F. Harrou, A. Saidi, Y. Sun and S. Khadraoui, "Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes," in IEEE Journal of Photovoltaics, vol. 11, no. 3, pp. 806-818, May 2021, doi: 10.1109/JPHOTOV.2021.3057169.

[5]. M. Aslam, S. -J. Lee, S. -H. Khang and S. Hong, "Two-Stage Attention Over LSTM With Bayesian Optimization for Day-Ahead Solar Power Forecasting," in IEEE Access, vol. 9, pp. 107387-107398, 2021, doi: 10.1109/ACCESS.2021.3100105.

[6]. F. Aziz, A. UlHaq, S. Ahmad, Y. Mahmoud, M. Jalal and U. Ali, "A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays," in IEEE Access, vol. 8, pp. 41889-41904, 2020, doi: 10.1109/ACCESS.2020.2977116.

[7]. C. -J. Huang and P. -H. Kuo, "Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting," in IEEE Access, vol. 7, pp. 74822-74834, 2019, doi: 10.1109/ACCESS.2019.2921238.

[8]. A. Guillén-Asensio, I. Sanz-Gorrachategui, A. Bono-Nuez, C. Bernal, J. M. Sanz-Alcaine and F. J. Perez-Cebolla, "Energy Shortage Failure Prediction in Photovoltaic Standalone Installations by Using Machine Learning Techniques," in IEEE Access, vol. 9, pp. 158660-158671, 2021, doi: 10.1109/ACCESS.2021.3129930.

[9]. K. Y. Yap, C. R. Sarimuthu and J. M. -Y. Lim, "Grid Integration of Solar Photovoltaic System Using Machine Learning-Based Virtual Inertia Synthetization in Synchronverter," in IEEE Access, vol.8,pp.49961-49976, 2020, doi: 10.1109/ACCESS.2020.2980187.

[10]. F. Li et al., "Detection and Diagnosis of Data Integrity Attacks in Solar Farms Based on Multilayer Long Short-Term Memory Network," in IEEE Transactions on Power Electronics, vol. 36, no. 3, pp. 2495-2498, March 2021, doi: 10.1109/TPEL.2020.3017935.

[11]. X. Li, W. Li, Q. Yang, W. Yan and A. Y. Zomaya, "An Unmanned Inspection System for Multiple Defects Detection in Photovoltaic Plants," in IEEE Journal of Photovoltaics, vol. 10,no. 2,pp.568-576, March 2020, doi: 10.1109/JPHOTOV.2019.2955183.

[12]. C. Kasburg and S. FrizzoStefenon, "Deep Learning for Photovoltaic Generation Forecast in Active Solar Trackers," in IEEE Latin America Transactions, vol. 17, no. 12, pp. 2013-2019, December 2019, doi: 10.1109/TLA.2019.9011546.

[13]. H. P. -C. Hwang, C. C. -Y. Ku and J. C. -C. Chan, "Detection of Malfunctioning Photovoltaic Modules Based on Machine Learning Algorithms," in IEEE Access, vol. 9, pp. 37210-37219, 2021, doi: 10.1109/ACCESS.2021.3063461.

[14]. S. Wen, C. Zhang, H. Lan, Y. Xu, Y. Tang and Y. Huang, "A Hybrid Ensemble Model for Interval Prediction of Solar Power Output in Ship Onboard Power Systems," in IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 14-24, Jan. 2021, doi: 10.1109/TSTE.2019.2963270.

[15]. H. Sheng, B. Ray, K. Chen and Y. Cheng, "Solar Power Forecasting Based on Domain Adaptive Learning," in IEEE Access, vol.8, pp. 198580-198590, 2020, doi: 10.1109/ACCESS.2020.3034100.

[16]. E. D. Obando, S. X. Carvajal and J. Pineda Agudelo, "Solar Radiation Prediction Using Machine Learning Techniques: A Review," in IEEE Latin America Transactions, vol. 17, no. 04, pp. 684-697, April 2019, doi: 10.1109/TLA.2019.8891934.

[17]. M. Alaraj, A. Kumar, I. Alsaidan, M. Rizwan and M. Jamil, "Energy Production Forecasting from Solar Photovoltaic Plants Based on Meteorological Parameters for Qassim Region, Saudi Arabia," in IEEE Access, vol. 9, pp. 83241-83251, 2021, doi: 10.1109/ACCESS.2021.3087345.

[18]. N. Khadka, A. Bista, B. Adhikari, A. Shrestha, D. Bista and B. Adhikary, "Current Practices of Solar Photovoltaic Panel Cleaning System and Future Prospects of Machine Learning Implementation," in IEEE Access, vol. 8, pp. 135948-135962, 2020, doi: 10.1109/ACCESS.2020.3011553.

[19]. U. K. Kalla, H. Kaushik, B. Singh and S. Kumar, "Adaptive Control of Voltage Source Converter Based Scheme for Power Quality Improved Grid-Interactive Solar PV–Battery System," in IEEE Transactions on Industry Applications, vol. 56, no. 1, pp. 787-799, Jan.-Feb. 2020, doi: 10.1109/TIA.2019.2947397.

[20]. K. Yan, Y. Du and Z. Ren, "MPPT Perturbation Optimization of Photovoltaic Power Systems Based on Solar Irradiance Data Classification," in IEEE Transactions on Sustainable Energy, vol. 10, no. 2, pp. 514-521, April 2019, doi: 10.1109/TSTE.2018.2834415.

[21]. L. Zhao, J. Li, Q. Li and F. Li, "A Federated Learning Framework for Detecting False Data Injection Attacks in Solar Farms," in IEEE Transactions on Power Electronics, vol. 37, no. 3, pp. 2496-2501, March 2022, doi: 10.1109/TPEL.2021.3114671.

[22]. Y. León-Ruiz, M. González-García, R. Alvarez-Salas, J. Cuevas-Tello and V. Cárdenas, "Fault Diagnosis Based on Machine Learning for the High Frequency Link of a Grid-Tied Photovoltaic Converter for a Wide Range of Irradiance Conditions," in IEEE Access, vol. 9, pp. 151209-151220, 2021, doi: 10.1109/ACCESS.2021.3126706.

[23]. L. Cheng, H. Zang, T. Ding, Z. Wei and G. Sun, "Multi-Meteorological-Factor-Based Graph Modeling for Photovoltaic Power Forecasting," in IEEE Transactions on Sustainable Energy, vol. 12, no. 3, pp. 1593-1603, July 2021, doi: 10.1109/TSTE.2021.3057521.

[24]. J. Huang, R. Wai and G. Yang, "Design of Hybrid Artificial Bee Colony Algorithm and Semi-Supervised Extreme Learning Machine for PV Fault Diagnoses by Considering Dust Impact," in IEEE Transactions on Power Electronics, vol. 35, no. 7, pp. 7086-7099, July 2020, doi: 10.1109/TPEL.2019.2956812.

[25]. H. Wagner-Mohnsen and P. P. Altermatt, "A Combined Numerical Modeling and Machine Learning Approach for Optimization of Mass-Produced Industrial Solar Cells," in IEEE Journal of Photovoltaics, vol. 10, no. 5, pp. 1441-1447, Sept. 2020, doi: 10.1109/JPHOTOV.2020.3004930.

[26]. B. Ray, R. Shah, M. R. Islam and S. Islam, "A New Data Driven Long-Term Solar Yield Analysis Model of Photovoltaic Power Plants," in IEEE Access, vol. 8, pp. 136223-136233, 2020, doi: 10.1109/ACCESS.2020.3011982.

[27]. A. Sharma et al., "Opposition-Based Tunicate Swarm Algorithm for Parameter Optimization of Solar Cells," in IEEE Access, vol. 9, pp. 125590-125602, 2021, doi: 10.1109/ACCESS.2021.3110849.