Development of Machine Learning based Technique for Solar Energy Estimation

A Thesis

Submitted in fulfillment of the requirements

for the award of the degree of

Doctor of Philosophy

in Engineering

Submitted by

Mr. Parveen Bhola

(Reg. no. 951504003)

Under the Supervision of

Dr. Saurabh Bhardwaj

(Associate Professor, EIED)

Electrical & Instrumentation Engineering Department

Thapar Institute of Engineering Technology

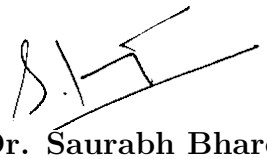
(Deemed to be University) Patiala-147004, Punjab, India

(September 2022)

CERTIFICATE

This is to certify that the thesis **"Development of Machine Learning based Technique for Solar Energy Estimation"** which is being submitted by **Parveen Bhola** to the Department of Electrical & Instrumentation Engineering, Thapar Institute of Engineering & Technology (Deemed to be University), Patiala for the award of the Degree of **Doctor of Philosophy**, is a record of bonafied research work carried out by him under my guidance and supervision and has full-filled the requirements for the submission of this thesis, which to by knowledge has reached the requisite standard.

The result embodied in this thesis have not been submitted in part or full to any other University and Institute for the award of any degree or diploma.



(Dr. Saurabh Bhardwaj)

Associate Professor, EIED

Thapar Institute of Engineering & Technology

(Deemed to be University), Patiala-147004, Punjab, India

Dedicated to my Parents Sh Sita Ram
& Smt Ramali Devi, Wife Shashi
and Kids Divyansh & Navya

ACKNOWLEDGMENTS

It is a great pleasure for me to express my respect and a deep sense of gratitude to my Ph.D. supervisor Dr. Saurabh Bhardwaj, Associate Professor, Electrical & Instrumentation Engineering Department, Thapar Institute of Engineering Technology, Patiala, for his wisdom, vision, expertise, guidance, enthusiastic involvement, and persistent encouragement during the planning and development of this research work. I also gratefully acknowledge his painstaking efforts in thoroughly going through and improving the manuscripts without which this work could not have been completed.

I am highly obliged to Prof. Padmakumar Nair, Director, Thapar Institute of Engineering Technology, Patiala, and Dr. Rafat Siddique, Senior Professor & Dean, Research and Sponsored Projects, and Dr. N. Tejo Prakash Professor & Associate Dean, Research and Sponsored Projects for the encouragement, support, and necessary facilities to carry out and complete this work on a steady course.

I am also thankful to present Chairman of the Doctoral Committee Dr. R.S. Kaler, Senior Professor & Head, Department of Electrical and Instrumentation Engineering, for the much-needed support throughout the work for providing all the facilities, help, and encouragement for carrying out the research work.

I am very thankful to Dr. Mandeep Singh, Professor, Department of Electrical & Instrumentation Engineering, Dr. M. D. Singh, Associate Professor, Department of Electrical & Instrumentation Engineering and Dr. Vinay Kumar, Associate Professor, Department of Electronics & Communication Engineering, for being the members of the Doctoral Committee and spending their valuable time reviewing and critically examining the work.

I also wish to express my deep sense of gratitude to Dr. Vikrant Sharma, Deputy Director (Technical), National Institute of Solar Energy, Gurugram, for providing me with the facility and the data required to carry out this work.

I would like to express my loving gratitude to my family members, especially to my Mother Smt. Ramali Devi and to my wife Mrs. Shashi for their patience, love, and encouragement during this journey.

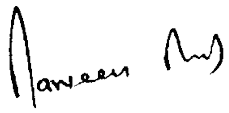
I wish to express my appreciation to my friends Tajinder, Gaurav, Maan Singh, Ro-

hit, Sikandar, Pankaj, Neeru, Jitender Virk, Vishal Chaudhry and thanks to research fellows at the department for their help and motivation throughout my research work. I also would like to express my deep and sincere thanks to my friends and all other persons whose names do not appear here for helping me either directly or indirectly in all even and odd times.

I am also thankful to the anonymous reviewers of my research publications. Their comments and suggestions were very helpful in shaping my research work.

I would also like to extend my special thanks to Mrs. Poonam, Arvind and other staff members of in EIE Department, for their timely help and cooperation, extended throughout the course of the investigation.

Finally, I am indebted and grateful to the Almighty for helping me in this endeavor.



(MR. PARVEEN BHOLA)

ABSTRACT

The Republic of India, which has the second-largest population in the world, has challenges in meeting its rising energy demands by reducing fossil fuels. Also, the energy produced by fossil fuels contributes to global warming as the energy sector is explicitly responsible for generating harmful substances during the production, distribution, and consumption of energy. Without posing any environmental risks, the Sun, a limitless energy source, can be used as an alternative to meet this rising need. Further, the advancement of technology in chemistry, material science, and solid-state physics improves the efficiency of photovoltaic (PV) modules, has resulted in various topologies with different performance characteristics, and added sun-fuelled plants to a portfolio of the electricity market. Despite their many advantages and relative popularity as a renewable energy source, Even the greatest solar panels eventually lose their effectiveness. Inspections are necessary to maintain cell performance levels and minimize financial losses since solar cells are susceptible to damage from weather-related incidents, temperature changes, and UV exposure over time.

How can real-time panel inspection be done in a way that is both economical and quick? This research work exploits the possibility of real-time estimation of solar power in PV systems. A new method was brought to light: Utilizing historical weather data, the Clustering-based Computation Rate (CCDR) calculates performance ratios and degradation rates. This method also allows for off-site inspection. Most methods on the market base do their calculations via on-site physical assessment of PV installations. These methods are not preferred for real-time degradation investigation because it is time-consuming, expensive, and labor-intensive. The suggested model provides a real-time estimation of the performance ratio. The degradation effect in terms of performance ratio is incorporated in estimation results. The degradation is calculated in real-time using the clustering-based technique without a physical on-site inspection, therefore, can be used for the real-time estimation of solar power.

Two important research goals are to maximize the power output from PV systems and further reduce economic losses. This study proposes a novel technique that uses a clustering-based technique to evaluate the degradation of PV panels with dif-

ferent topologies. The proposed method calculates the degradation in output solar power in terms of PR for panels with three different topologies, namely, amorphous silicon (a-Si), polycrystalline silicon (p-Si), and heterojunction with an intrinsic thin layer (HIT), over a period of three years. The degradation rate produced by a-Si technology was the lowest, and it was highest for HIT technology. The degradation is calculated in real-time using the clustering-based technique is used further to fine-tune the estimation of solar power. Initially, the power is estimated using the Support Vector Regression (SVR) model with the meteorological parameters. The estimation is further fine-tuned in sync with the degradation rate. The model is validated on the actual data (Meteorological parameters and Solar power) procured from the solar plant. After refinement, the estimation results show significant improvement in terms of statistical measures.

Contents

List of Figures

List of Tables

List of Acronyms/Abbreviations

PV	Photovoltaic
ANN	Artifical Neural Network
ANFIS	Adaptive Neuro Fuzzy Inference System
AR	Autoregressive
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
DNI	Direct Normal Irradiance
DHI	Diffuse Horizontal Irradiance
GHI	Global Horizotal Irradiance
DWT	Discrete Wavelet Transform
FFNN	Feed Forward Neural Network
GSI	Global Solar Irradiance
GFM	Generalized Fuzzy Model
MA	Moving Average
SVM	Support Vector Machine
SVR	Support Vector Regression
MLFFNN	Multilayer Feedforward Neural Network
RNN	Recurrent Neural Network
NWP	Numerical Weather Prediction
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
MAPE	Mean Absolute Percentage Error

RMSE	Root Mean Square Error
NRMSE	Normalised Root Mean Square Error
PR	Performance Ratio
FY	Final Yield
CUF	Capacity Utilization Factor
a-Si	Amorphous Silicon
p-SI	Poly Silicon
HIT	Hetero junction with Thin Layer
CSD	Classical Seasonal Decomposition
YoY	Year on Year basics

List of Symbols

Y_R	Reference Yield
Y_F	Final Yield
T_{Avg}	Temperature Average
H_o	Solar Radiation
N	Number of Days
R_H	Relative Humidity
P_{Avg}	Pressure Average
D_{Pt}	Dew Point
U	Wind Speed
S_t	Sunshine Hours
R^2	Coefficient of Determination
P_m	Rated Power
E_{ac}	Actual Energy produced
X,	Normalized Data
$L_\epsilon(d, y)$	Loss Function for SVR
ξ	Slack Variables
ξ,	Slack Variables
α	Lagrangian Multiplier
α,	Lagrangian Multiplier

Chapter 1

Introduction

Chapter 1 introduces the motivation for the development of machine learning-based technique for solar energy estimation. Afterward, the thesis contribution and literature survey on solar energy techniques are presented, and finally, the thesis organization and objectives are presented.

1.1 Background

The estimation of solar energy has received scientific attention in recent years, thanks to the advancement and widespread adoption/application of solar-based technologies. Researchers have developed multiple techniques for estimating solar energy based on existing observed values and their dependencies on various meteorological parameters. In recent years, machine learning and deep learning algorithms have been explored for accurate solar energy forecasts. This research provides an accurate solar power estimate using two powerful machine learning techniques, i.e., Support Vector Regression (SVR) and K-Means Clustering. Another important aspect of this study is considering the effect of degradation induced in PV panels. The K-means clustering technique, an unsupervised machine learning method, provides the degradation information. This information helps us to estimate solar power accurately. Therefore, this research exploits the possibility of real-time estimation of solar power for PV panels.

A solar cell is a device that transforms photons in solar rays to direct current (DC) and voltage. It is the heart of the PV system, and the associated technology is known as solar Photovoltaic (PV). When light strikes the solar cell, it absorbs the light and transforms it into the recombination of electron-hole pairs. A silicon solar cell's operational process is depicted in Figure 1.1. The efficiency, size, and sunlight that strikes a PV cell's surface affect how much current it can produce. A commercial PV cell with a surface area of around 25 square inches will generate approximately

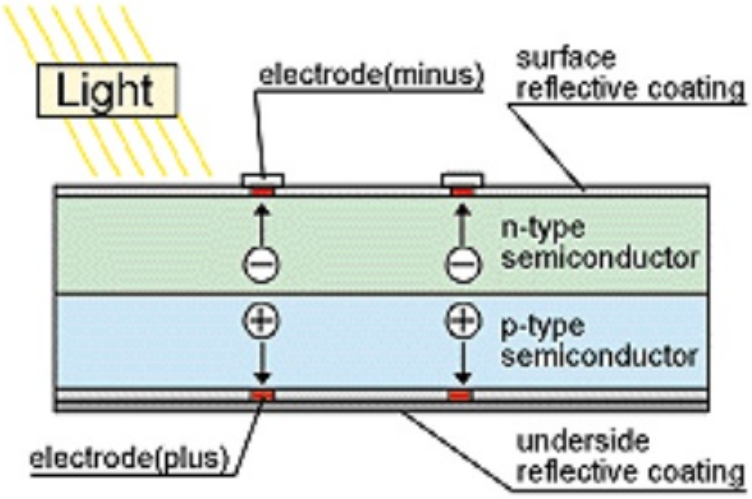


Figure 1.1: Solar Cell Working

2 watts of peak power when exposed to the most sunlight. Strings or arrays of solar cells are created. Additional series and parallel connections are made between the arrays to produce the required power. This work's main area of focus is estimating solar power in PV technology. The PV panel technology considered in this work is shown in Figure 1.2.

Figure 1.2: Mono, Poly Crystalline and HIT Silicon Cell

1.2 Motivation for the Research

The main concern in any Nation's development is energy. With the abatement of fossil fuels, we have to look up other alternatives/sources of energy. Sun is a huge source of energy. Therefore, sun fuelled plants are added to the portfolio of the electricity market over the world. Also, the energy produced by fossil fuels contributes

to global warming as the energy sector is explicitly responsible for generating harmful substances during the production, distribution, and consumption of energy. The advancement of technology in chemistry, material science, and solid-state physics improves the efficiency of PV modules. It has resulted in various topologies with different performance characteristics that have boasted solar power. Implementation of PV technology is significantly growing. Therefore, researchers have shown a keen interest in the performance monitoring of PV systems. Resistant to degradation, a decreasing trend in power output over time is a crucial performance parameter. Since the installation cost of PV systems is higher than that of traditional systems, an estimation of the energy output at the location where the system is to be installed is also necessary for the design of a solar PV system. It is helpful to be aware of the degradation rate to predict energy output accurately. Additionally, active monitoring of PV system degradation helps determine underperforming systems and reduces financial losses brought on by operational issues. Considering all the above factors, solar estimation is a great area to work on.

1.3 Thesis Contribution

This thesis contributes to the possibility of real-time estimation of solar power from a solar PV system. Clustering-based computation Rate (CCDR), an innovative technique for calculating PV panel performance ratios and degradation rates, has been introduced. Off-site inspection is another feature of this approach. Most of the methods available do on-site physical assessments of PV systems. These methods cannot be utilized for real-time degradation investigation because they are time-consuming, expensive, and labor-intensive. Following are the main contributions of the proposed work:

- Development of the Machine Learning based technique to estimate the Solar radiation and Power.
- Development of a new clustering-based approach to calculate the degradation of P.V. panels in real-time. Amorphous silicon (a-Si), poly-crystalline silicon (p-Si), and HIT technology panel topologies are considered for validation.

- To fine-tune the estimation of solar power by incorporating the real-time degradation in PV panels using the developed technique.
- Validation of the proposed technique on the real data procured from the solar plant.

1.4 Need for Solar Energy Estimation

As solar energy primarily depends upon solar radiation, therefore the first and most essential task is the estimation of global solar irradiance. Due to rising global demand and emphasis on solar energy as a source of total power, accurate estimation of solar energy is required. Additionally, to have the proper mix of power from renewable and fossil fuels, energy planners and electric utilities need an accurate estimate. Its exact estimate is also important for designing a solar plant. Additionally, accurate forecasts will aid with maintenance, energy storage, and trading of electricity. They will also prevent unnecessary fuel usage and emergency electricity purchases. Depending upon the application and the time horizon, the solar radiation estimation can be categorized as per Figure 1.3.

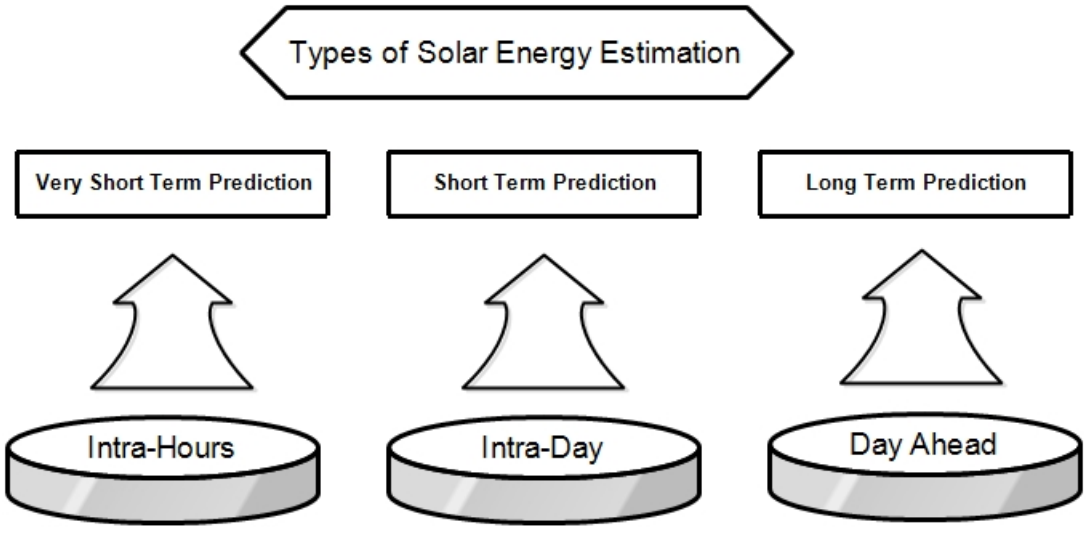


Figure 1.3: Types of Solar Energy Estimation

- Intra-Hour: Very short-term prediction deals with the time horizon of 15 minutes to 2 hours ahead. For the purpose of tracking battery state in real-time, very short-term forecasting is crucial.

- Intra-Day: Short-term prediction deals in the time horizon of 1 hour to 6 hours ahead. For operations involving decision-making, such as unit commitment, etc., short-term forecasting is essential.

- Day-Ahead: Long-term prediction deals with the time horizon of days or weeks ahead. In order to plan network operations, long-term forecasting is helpful. Day Ahead predictions are necessary for operational planning, source switching, programming backup, short-term power purchases, and the scheduling of reserve utilisation.

1.5 Literature Survey

This research focuses on developing machine-learning-based methods to estimate the PV power output. This section provides a detailed literature review to analyze the previous methods for estimating solar energy. This survey aims to understand the current techniques better so that a better solar energy forecasting method may be developed using machine learning methods. The methods reported in the literature for solar energy estimation can be largely divided into the following two categories:

- Direct Methods

- Indirect Methods

1.5.1 Direct Method

The direct group of strategies estimates the power output of PV systems directly without an initial estimate of solar irradiance as required in the case of indirect methods. The primary data sources are previous PV power data and additional data like historical weather data of the PV plant location. This information is less sophisticated and readily available than that needed by indirect methods. These methods consist mainly the statistical techniques, machine learning algorithms, and hybrid models. The relationship between solar irradiation and other variables like temperature, pressure, wind speed, etc., is predicted using these approaches using meteorological data.

1.5.1.1 Statistical Method

These methods mainly consist of the Persistence Model, Autoregressive Moving Average (ARMA), and its expansion, Autoregressive Integrated Moving Average (ARIMA), and Exponential Smoothing (ES) models. They are generally applied for short-term estimation (within day or day ahead). In comparison to indirect methods, prediction models can be used directly to estimate PV power outputs without needing particular technical expertise. The persistence model is the superficial model used as the reference model. Any model developed for the estimation of solar energy for short-term prediction must outperform the persistence model. The persistence model is only used for comparing the merits of developed techniques by the researchers, this is an inaccurate method for prediction for more than 1 hour ahead of estimation, and its accuracy decreases with cloudiness changes from the current state. Autoregressive Moving Average (ARMA) is another important tool used for estimation. The data generated or obtained sequentially, known as time-series data, is required for the ARMA model [1]. It combines the autoregressive (AR) approach and the moving average (MA) technique. The model is named the ARMA (p, q), where p represents the autoregressive measure, and q shows the moving average part. Box-Jenkins performs the ARMA to extract statistical information from the time-series data to predict future values, but the condition is that the series should be stationary. The research article [2] examined six non-exogenous input forecasting models. They evaluated the accuracy of ARIMA against the persistent method, k-NN, NN, and a NN merged with Genetic Algorithms (GA-NN) using data spanning eight months. Although GA-NN outperformed the other techniques in the comparison, ARIMA also achieved good accuracy. X. G. Agoua et al. [3] built a statistical model, the spatiotemporal technique, to estimate power output from a few minutes to 6 hours in advance. They developed an innovative stationarization approach to address the issue of time series non-stationarity. The findings indicate that pre-processing was helpful since it resulted in significant performance instead of using actual data. They also mentioned that integrating meteorological factors like wind power enhances the spatiotemporal model. Compared to a random forest, AR, and persistence model, the suggested technique was 20% more accurate. Researchers in New Zealand used another statistical approach to estimate a 24-hour ahead global irradiation using dif-

ferent techniques but focusing on nonlinear autoregressive recurrent neural networks having exogenous inputs (NARX) mentioned in reference [4]. The model was trained and tested using hourly time series data for nine historic meteorological parameters acquired over three years. The study included reference persistence, ANN-based Multilayer Perceptron (MLP), statistical (ARMA) technique, and NARX-based forecasts. The hourly predictions were compared to measured values in terms of the root mean squared error (RMSE) for the NARX method, and it was found to be 0.243 MJ/m2, better than the MLP, ARMA, and persistence approaches, respectively. V. Prema et al. [5] performed the short-term prediction (one day ahead) of solar irradiance in Microsoft Excel and R software environment using moving average (ARMA), exponential smoothing, and decomposition model. ARMA performs better for the lower values of estimation. The decomposition method performs well for significant duration estimates considering two months span. The result was compared in terms of MAPE, which was observed at 9.28% for the decomposition method. The error increases for cloudy weather but remains in the permissible zone. The research article [6] presents a multiplicative ARMA model for generating instantaneous global irradiation series. The data set consisted of five-minute global irradiance data collected at a radiometric station in Córdoba, Spain, from 1994 to 1997. The seasonal fluctuation and annual periodicity are taken into consideration during estimation. The proposed approach incorporates training an AR to the data. When estimated solar radiation levels are matched to the actual data, it produces the best results.

In summary, statistical methods and prediction is based upon the requirement of the time series data of meteorological variables. Further, assuming that a linear process generates the data, it is not true as all the practical problems develop nonlinear data. The statistical method does not allow the ability to adapt to non-linearity and approximate complex relationships. Hence, the researchers adopted machine-learning methods like ANN, Fuzzy Logic, SVM, etc.

1.5.1.2 Machine Learning Method

Neural Networks (NNs), k-means clustering, and Support Vector Machines techniques are frequently reported in the literature on solar energy estimation. The primary requirement for these models is the data set, and they do not need the expertise

in electric power as required in the case of indirect methods. Another advantage of these techniques is adjusting the forecasting time [7]. The methods discussed in previous sections are suitable for very short-term or short-term estimates. At the same time, the forecast horizon of machine learning techniques is adjustable, varying from intra-hours to day ahead forecast [8].

1.5.1.3 Neural Networks (NNs)

For solar energy estimation, NNs are the most employed method. NNs can easily handle complex nonlinear problems, but careful parameter selection is necessary, along with the NN structure and training procedure [9–15].

J. C. Lam et al. [16] studied the MLFFNN to estimate daily global solar radiation in China. Latitude, longitude, altitude, day number, temperature, and sunshine ratio are the inputs provided to the neural network. The created MLFFNN's model accuracy is assessed, and it is found to have an RMSE value of (9.1–20.5%) and MBE values of 16.9% to 18.6%.

F. Almonacid et al. [17] used solar irradiance and temperature as the two inputs of the neural network model to estimate the one-hour ahead power output for a solar plant at the University of Jaen. The suggested approach was tested using a linear regression model, which verifies the results between the estimated and actual values. The proposed model showed a correlation coefficient close to one, and a root mean square error of 3.38 percent.

Dahmani et al. [18] prepared an NNs model to predict the tilted global solar irradiance based on horizontal data collected in Algeria. The model utilizes horizontal global extraterrestrial irradiation at 5-minute intervals and zenith angle, azimuth angle, and declination as inputs. It is tested on two years data set. The results produced by the model were satisfactory, with the best value of RMSE 8.82 percent. Teo et al. [19] also made a similar observation. They suggested extreme learning machine technique implies a NN model to estimate solar power directly. They evaluated the model's performance using three data sets. They found that changing input factors and increasing the training data set could enhance performance substantially. However, another decision was reached. in [20]. G. Notton et al. suggested three different NN models, tested with data set collected in France from a PV plant over five years

of data at 10-minute intervals. The input parameters for the first two NN models were declination, zenith angle, time, 10-min extraterrestrial irradiation, and 10-min extraterrestrial horizontal irradiation. The inclination angle was considered an extra input to the third NN model. The results revealed that removing one variable from input improves the results. The errors were reduced by 9% (RMSE) and 5.5%, Relative Mean Absolute Error (RMAE).

RNN method for solar energy estimation is applied in [21]. The RNN model is preferred for sequence prediction and works well in solar estimation. To train the RNN smoothly, Yona et al. apply fuzzy algorithms to obtain insolation estimation. The suggested model outperformed the other models in the comparison (a model based solely on fuzzy technique, a persistence model, and a feed-forward NN), with an MAE of 0.1327kW.

Rehman et al. [22] used an ANN method to estimate global solar radiation in Abha, Saudi Arabia using the two input parameters, relative humidity and air temperature. Based on these two variables, the ANN model estimated global solar radiation effectively with an absolute mean percentage error of 4.49% compared to other standard methods. However, Ozgoren et al. [23] used the multi-nonlinear regression (NLR) method to develop an ANN model to predict monthly global solar radiation over Turkey using several input variables, including latitude, longitude, altitude, month, monthly minimum atmospheric temperature, atmospheric pressure, vapour pressure, maximum atmospheric temperature, relative humidity, wind speed, rainfall, cloudiness, mean atmospheric temperature, soil temperature, and sunshine duration. The proposed technique was used to identify more influential inputs. Their findings revealed that the ANN model's projected reasonably statistical result verified with the measured values. The MAPE for the testing data set was 5.34%, and the coefficient of correlation (R) value was around 0.9936. In a similar study in the Mediterranean region of Anatolia in Turkey, Koca et al. [24] conducted a global solar radiation estimation study using the ANN approach. They looked at the impact of six distinct input parameter combinations on the ANN technique's performance. Their findings show that ANN accurately estimated global solar radiation at a part in the studied region with an RMSE of 3.8%.

Azimi et al. [25] suggested a new hybrid technique for estimating solar radiation

that combines a clustering approach with NNs. They first divided the time series data for solar power to find unusual patterns and outliers into distinct sets using a transformation-based k-means method. The final predictions are then trained on the clustered data by neural networks (NNs). The created methodology was compared to other statistical models, such as ES and ARIMA, and a persistence model and was shown to be the most accurate.

1.5.1.4 K-Means Clustering

Clustering's primary goal is to organize things into groups or clusters so that data in the same cluster are alike while data in other clusters are different. The k-means clustering technique, which helps categorize any given data, is one such clustering technique that has attracted the research community. This method of learning is unsupervised. K-means has several uses in several fields of study, particularly in predicting renewable energy. This method helps to manage data groups separately, allowing one to understand the data better and improve estimation quality.

Matteri et al. [26] suggested a fusion of K-means clustering algorithms and ANNs for day-ahead PV power forecast. The clustering algorithm split the dataset into different day classes with identical weather conditions. Then, a separate ANN is created for each cluster to improve the estimation result. The most significant improvement was a 7.9% reduction in RMSE error. In a similar study in china, M. Li et al. [27], the K-means clustering technique splits the input data into five zones from 90 different sites. The model for estimating solar radiation for each zone was developed using geographical characteristics (longitude, altitude, and latitude). The model performance was investigated using the MBE, MPE, RMSE, and coefficient of residual mass (CRM). The models had satisfactory accuracy compared to measured and projected daily solar radiation.

TR Ayodele et al. [28] applied a k-means clustering approach to split the seasonal meteorological data into distinct clusters that were not affected by the seasons. This arranges the clustered data into four different sky conditions batches and then is applied to SVR models for estimation purposes. The model's accuracy was tested with six years of meteorological data (2010–2015) to forecast daily global solar radiation for the coming years (2016–2017) collected from Ibadan, South Africa. The hybrid

model outperforms a standard method like the ANN and ARMA model with RRMSE of 2.6515%, MAPE of 1.7928% for 2016 and RRMSE of 2.7498% and MAPE of 1.795% for the year 2017. Pedro F. et al. [29] reported a similar study in Spain. The authors estimate next-day hourly solar radiation values using decision trees, ANN, and SVM with an MAE of 15.2 percent for one of the input data sets and 16.7 percent for the other input parameter set.

The article in [30] presents a simple method for predicting global radiation in areas where only surface meteorological measurements are available. The data obtained from 1042 surface weather stations in China was used to identify nine solar radiation zones using k-mean clustering. The zone models made for prediction are reasonably accurate, with RMSE percent less than 20% in about 89% of stations. Ji Wu et al. [31] conducted a similar study in Singapore, where the solar radiation series data is first split into smaller subsets to obtain the different clusters of similar patterns in the data. A suitable prediction model trained for each set outperforms in comparison with standard model ARMA in terms of RMSE.

Hamza Ali-Ou-Salah et al. [32] presented a hybrid approach for forecasting one h-ahead global solar radiation utilizing a seasonal clustering algorithm and ANN. The fuzzy c-means (FCM) method finds clusters for three years of experimental data with different seasons based on Portugal's solar and meteorological characteristics. Different training subsets were created from the meteorological dataset obtained from clustering. ANN model was developed for each subgroup to estimate hourly global solar radiation. The results are compared to statistical measures like RMSE, which was reported at 106.4 W/m^2.

1.5.1.5 Support Vector Regression

Besides the NNs, SVM [33] can be another cutting-edge machine-learning tool extensively employed for forecasting solar power by the research community. However, for the estimation tasks, the Support Vector Regression (SVR) variant of the SVM is used [34–36].

Rana et al. suggested a 2D-interval estimation technique based on SVR in [34], directly estimating 2D-interval solar power output from old power and weather data. For two years, the model was tested using Australian PV data. Compared to some

benchmarks and approaches utilized for comparisons, such as NN2D and two persistence models, their findings prove the SVR2D offered the best accurate forecasts.

Ramli et al. [35] used data from Jeddah and Qassim in Saudi Arabia to compare SVM and NN for solar irradiance estimations. They employed direct diffuse and global solar irradiance as input and computed the models' MRE, correlation coefficient, RMSE, and computing speed. The SVM approach gave higher accuracy and more stable calculation, with MRE = 0.33 and 0.51 for both cities, respectively, and a fast prediction speed of 2.15s.

K. Mohammadi et al. [36] proposed a new hybrid technique integrating the SVM with Wavelet Transform (WT) to estimate horizontal global solar radiation. Predictions for an Iranian coastal city are made on daily and monthly mean levels. It is compared to various current techniques to show the effectiveness and feasibility of the proposed SVM–WT method. For the best SVM-WT model, the research findings for MAPE, MABE, RMSE, NRMSE and coefficient of determination (R2) for daily estimation are, in order, 6.9996 percent, 0.8405 MJ/m^2, 1.4245 MJ/m^2, 7.9467 percent, and 0.9086.

Shi et al. [37] presented a study to estimate day ahead solar power using the SVM and weather categorization characteristics. Based on the local weather report, correlation analysis was used to separate the time series of PV power output data into four categories. Four SVM models are constructed based on the features of the data samples, which include the maximum, minimum, and mean temperatures for various weather classifications. The maximum accuracy is shown in terms of RMSE 1.57MW, whereas the least value of RMSE reported was 2.52MW.

Mellit et al. [38] suggested an LS-SVM model produce short-term predictions for meteorological time series. As input parameters, they took into account the sun irradiance, wind speed, temperature, relative humidity, atmospheric pressure, and wind direction. The LS-SVM model was more accurate than the NN models in prediction when the SVR model was compared to four distinct NN models (MLP, RBF, RNN, and PNN).

Wolff et al. used SVR techniques for solar power prediction for 15-min and 5-h forward time horizons [39]. This technique was a replacement for estimation models like the NWP. Their findings revealed that SVR offered better results for 1-hour forecasts,

while NWP-based models were good for 3-hour forecasts. The researchers believed that integrating the results of various prediction algorithms could enhance accuracy even more.

Ekici [40] suggested an LS-SVM hybrid approach with an RBF kernel function to predict solar energy for the next day. The proposed model's input parameters included the daily mean and maximum temperatures, the length of the sun's shine, and the daily solar radiation's historical data. The results demonstrated the success and suitability of the suggested model for the work.

Y. Liu et al. [41] proposed a new whale algorithm to improve the SVM method results. The feature of the model is a requirement for less training data. It adapts symmetrically to multiple weather forecast situations and has excellent estimation accuracy in varied weather scenarios. The input parameters are relative humidity, ambient temperature, and light intensity. The improved whale algorithm refined the SVM model to boost prediction accuracy. Compared to other standard methodologies, the suggested IMWOA-SVM model produced excellent estimates for a sunny and cloudy day, notably in cloudy conditions, with MSE 0.257, RMSE 0.507, and MAE 0.331.

Chan et al. [42] created a new least-squares support vector regression with a Google (LSSVR-G) technique that can estimate PV power in Taiwan. This study combines a Google application programming interface (API), LSSVR, and a GA. Compared to MAPE 19.25 and MSE 340.46, the results confirm that the LSSVR-G model is a good choice for projecting renewable power generation.

B. Mohammadi et al. [43] attempted to predict solar radiation with a new hybrid technique combining support vector regression (SVR) with the Krill-Herd algorithm (SVR-KHA) model. The solar radiation values for eight stations were predicted by applying the SVR-KHA algorithm approach based on SVR and implementing nearby station data to achieve this objective. Results demonstrate the performance of SVR-KHA was better and had minor errors compared to traditional SVR. Further, this model's RMSE, MAPE, and R2 values were 1.98 MJ/m2/day, 7.4 percent, and 0.93, respectively.

The authors proposed an essential enhancement for regression estimation in SVR over the widely used 'ϵ- insensitive loss function' [44]. They developed the 'ϵ-penalty loss function' for the 'ϵ-Penalty Support Vector Regression' (ϵ-PSVR) model, which is

superior to existing loss functions utilized in various SVR models for general noise distribution. The proposed SVR models have a more excellent generalization capability than existing SVR models.

Apart from using SVR in solar, researchers also deployed it for other applications. One of the applications is rain predictions at two rain gauge locations in northwest Iran. The author used a unique approach to create a hybrid regression model for 1-month-ahead rainfall forecasting. This method combines the SVR with the firefly algorithm (FFA) to produce accurate rainfall forecasts [45]. Another area that was explored is load forecasting. The author developed a hybrid model using SVR and PSO to forecast the load from the data received from the Australian electricity market [46]. Another area that was explored is load forecasting. The author developed a hybrid model using SVR and PSO to forecast the load from the data received from the Australian electricity market [47]. A similar study for load forecasting is reported in [48, 49].

1.5.2 Indirect Method

The energy produced by any solar PV plant mainly depends upon the solar irradiance of the site location. The indirect method estimates the solar irradiance of plant location, where it is generally unavailable due to the high cost of instruments. Therefore, the indirect method combines NWP algorithms and satellite/digital images to estimate solar irradiance. After estimating the solar irradiance, this information is used to estimate output using the parameters of PV plant. That is the reason it comes in the indirect methods. Finding the solar irradiance indirectly requires complex meteorological data like cloud cover movement, sun angle changes, etc., which are analyzed using NWP and satellite images to predict solar irradiance. Most NWP models use complex equations, which need area expertise to describe the radiation process and atmospheric fluctuations [50–52]. Cornaro et al. [53] highlighted the main advantage and drawbacks of NWP. The main advantage is the deterministic physical approach of the model, which means the solar irradiance of the plant location can be easily calculated where it is unavailable. However, the NWP method is constrained by the nonlinearity of the introduced equations and the poor spatial resolution. NWP model's spatial resolution ranges from 100 km to a few km which is too broad com-

pared to the size of a PV plant. NWP models do not precisely estimate small-scale features due to their small resolutions. NWP models' accuracy is subject to the readiness of meteorological information. NWP perform better when used for short-term forecasting, so they are not suitable for the long term, as in the case of this study. Another Indirect approach is to capture cloud images through satellite or ground cameras. The majority of digital camera models available develop the hemispheric image of the sky with cloud movement used to estimate the irradiance. The cloud movement through the cloud motion vector is used to forecast cloud cover, irradiance, and solar power in the short-term estimation. How the cameras are set up significantly impacts the tracking of the cloud and their detection strategies. Chow et al. [54] presented a ground-based sky imaging technology. They took sky photos every 30 seconds and used a clear sky model and sunshine characteristics to determine the sky cover. They used a two-dimensional cloud map they developed from co-ordinately transforming the sky cover to estimate the surface cloud shadows. Cloud speed and forecast horizon were the two crucial parameters influencing estimation accuracy. According to the findings, the estimation error was decreased to 50%-60% of the persistence model error in the 30s projections. In a similar study presented in [55], where the images are taken from the satellite. They used numerous sky images to build a short-term solar irradiance prediction model for new 3D cloud identification and tracking. They employed a classifier and the output cloud mask to identify clouds at the pixel level. Then, using the images from the various sky imagers, they calculated each cloud layer's block-wise base height and movement, which could merge into extensive views for the solar forecast. For all irradiance estimates between 1 and 15 minutes intervals, the suggested model outperformed the persistence model by at least 36%. Another method used a hybrid approach combining the satellite image and neural networks presented by Marquez et al. [56]. They produced GHI forecasts at 30, 60, 90, and 120-minute temporal horizons. The parameters for cloud fraction were calculated and applied to NN as inputs. The suggested method surpassed the persistence approach by 5-19% for single time-step estimates and nearly 10-25% for multi-step estimates.
In Summary, Indirect approaches are significantly dependent on the NWP model, cloud motion vector, and images (taken from ground and satellite) and require the skills (complex equations) to apply the model for estimating solar energy. Therefore,

the use of these methods makes it less popular.

1.5.3 Summary

The modern approaches for solar power prediction were introduced in the previous sections. These methods are separated into two categories: Indirect and Direct methods. When putting these ideas into practice, each group has its unique set of strengths and limitations that must be considered.

The indirect method is based upon the meteorological parameters, which is helpful for solar power estimation, e.g., the movement of clouds directly influences the PV panels' output, and the temperature of surroundings affects the rate of conversion in PV panels. As a result, meteorological techniques using NWP data and satellite pictures are employed globally. The former incorporates global or local meteorological data that can affect the variability of PV power output, whereas the latter is better at estimating cloud movement. However, the meteorological models based on indirect assessment also have some limitations. As previously stated, their performance is based on reliable weather estimates, which may or may not be present for the PV plant's site. The absence of monitoring variables like wind speed and cloud cover index, or the inaccuracy of forecasts for those variables, can considerably reduce predicting accuracy. Therefore, indirect methods are not preferred for solar power forecasts in practice. Another disadvantage is not able to perform adequately for short or extremely short time horizons [54] [56]. This happens because the essential tools for indirect methods, such as NWP or satellite images, are better for intra-hours and intra-day jobs. When clouds form and dissipate quickly, forecasts are affected for hours ahead of time. Due to their short forecasting horizon limit, they are less suited to tasks needing longer forecasting horizons, like those of 24 or 48 hours in the future.

The direct method mainly relies on the meteorological parameters and the power data, which are easily available, and not dependent upon the weather forecast as in the case of the indirect method. Our discussion mainly concerns the direct methods comprising Statistical and Machine Learning methods. The statistical group uses algorithms like AR, MA, ARMA, and ARIMA, which can directly predict PV power/radiation output [1] [57]. They are suitable for short-term estimation goals. However, their forecast accuracy suffers when horizons are extended to a day ahead. This limits the

use of statistical approaches in reality. Another challenge with statistical models is typically oriented on the time series data. They are not preferred to deal with sudden changes, like in weather conditions, particularly over a short period, where indirect methods excel.

Machine learning techniques are also commonly used to forecast PV power output. They also do not need much meteorology and solar power engineering expertise, so they are accessible to a broader range of data scientists. Another advantage of machining learning methods over meteorological and statistical techniques is the estimation horizon time flexibility. Machine learning methods can handle short-term (intra-hour), long-term (day-ahead), and even month-ahead forecasts, giving PV plants more flexibility in running their operations[[8]. Machine learning approaches are also versatile, as they may predict PV power output in direct and indirect methods. When meteorological data is unavailable, they can still do estimation using solar power data, similar to statistical models. This extends their usability because reliable weather data for the PV plant's location is not always available. Machine learning models may include more input features, which may improve the models' performance. However, some aspects of prior machine learning techniques for solar power estimation might be further explored and enhanced. There is still space for improvement in the techniques, which have employed clustering approaches to classify days using weather data and develop a distinct prediction PV model for each cluster [26–29].

This thesis explores the application of machine learning techniques to estimate PV output for the coming day directly. At the same time, we focus on two primary issues that have not been well addressed in earlier research:

Clustering-based computation of degradation rate (CCDR): Previous research used clustering approaches for solar power forecasting tasks but did not account for PV panel degradation. We propose to conduct a thorough investigation of these constraints. This work is explained in Chapter 3.

Accuracy improvement is solar estimation using Real-Time degradation computation: Accuracy of solar power estimation is refined by considering the degradation information as one of the input parameters of the machine learning model. The accuracy is compared before and after refinement in various statistical measures. This work is discussed in the chapter 6.

1.6 Research Gap

Research is a continuous process, evolving novel techniques and concepts for the betterment of society. Researchers have made numerous efforts so far in solar energy estimation, but still, there is scope for improvement. After going through the literature, the following research gaps are identified.

Various approaches have been attempted for solar power estimation using satellite images, cloud motion vectors, and weather data. Numerical Weather Prediction (N.W.P.) is another approach used for long-term prediction, is still in its infancy, and is not in practice for predicting solar radiation. Therefore, hybrid models using soft computing techniques can be used for prediction, which is more suitable. As per the available literature, no author has highlighted the issue of a decline in power rate in P.V. technology over time in estimation. The efficiency of solar power is dependent on the PV module, Inverter, and charge storage devices used; this decline in power rate over the period is known as the degradation rate. To accurately predict solar power, it should be taken into consideration during the estimation of solar Energy. Based on the aforementioned problem definitions, the ultimate goal of my research is to develop a machine learning-based technique for solar energy estimation, including the degradation information.

1.7 Objectives

The following objectives have been proposed in this research work.

- Pre-processing and analysis of meteorological data.
- To develop a machine learning-based model for radiation and P.V. array energy output estimation.
- To modify/develop an algorithm to include the degradation effect for various P.V. technology and re-estimate the energy output.
- To validate radiation and energy output estimation model for different geographical regions.

1.8 Organization of Thesis

The research work presented in the thesis is organized and structured in the form of seven chapters, which are briefly described as follows:

i) **Chapter 1** The key sections of this chapter are an introduction, a literature survey of solar energy estimation techniques, research gaps, and objectives of the research work.

ii) **Chapter 2** In this chapter, the methodology used to find the solar energy estimation technique will be discussed in comparison with other available techniques.

iii) **Chapter 3** This chapter is dedicated to the degradation part. Here, the technique used for degradation calculation will be discussed along with the other standard methods available for degradation computation.

iv) **Chapter 4** This section is dedicated to the data and site location used for the research work. The pre-processing and conversion of data techniques are discussed here.

v) **Chapter 5** This chapter elaborates on the experimental results obtained for solar radiation, solar power, and degradation computation via the CCDR method.

vi) **Chapter 6** This chapter elaborates on the real-time estimation of solar power using the degradation information obtained in chapter 4 for estimating the real-time solar power.

vii) **Chapter 7** The conclusion and future scope of this research work have been discussed in this chapter.

Chapter 2

Methodology for Solar Energy Estimation

This chapter is dedicated to the literature survey on estimation. Here, we have tried to find the researcher's various techniques for solar energy estimation. Afterward, one of the popular Machine Learning, Support Vector Regression (SVR), is elaborated on and applied as an estimation tool for the present research work. The various performance measuring parameters are also discussed in the last.

2.1 Historical Background of Estimation Techniques

Solar energy is obtained in the form of solar radiation. A large amount of energy is released as electromagnetic radiation comes from the sun. Solar energy provided is interpreted as the amount of solar radiation energy incident on the earth's surface per unit area per unit time. Global Solar Irradiance (GSI) and solar irradiance on the horizontal surface from direct sunbeams and diffuse sky radiation are two ways it can be detected. When radiation does not reach a surface directly from the Sun due to atmospheric molecules and particles scattering it, it is referred to as diffuse horizontal irradiance or DHI. Another terms used are Direct Solar Irradiance(DSI) or Direct Normal Irradiance(DNI) on a surface held perpendicular to the sun rays, and diffuse sky radiation is obstructed. Extra-terrestrial Radiation is the intensity of solar radiation at the top of the earth's atmosphere incident normal to the sun. It varies throughout the year due to the earth's elliptical orbit with a varying predictable distance between them. Therefore, the first and most crucial task for estimating solar energy is to estimate solar radiation.

Diffuse solar energy is usually measured by pyranometers, or actinography, while direct beam solar radiation is measured by a pyrheliometer. These measuring devices are usually installed at selected sites in specific regions and it is not feasible to install them at many sites due to the high cost of these devices. To create solar energy models

that explain the mathematical relationships between solar energy and meteorological variables such as ambient temperature, humidity, and sunshine ratio observed solar energy values can be used. At locations without solar energy measuring equipment deployed, these models can be utilised in the future to forecast direct and diffuse solar energy using past meteorological data.

2.2 Literature Survey for Estimation Techniques

Our country, India, receives the sun's radiation in abundance. The solar irradiance in different geographical locations varies from 4 - 7 kWh/m^2 per day with 2300 - 3200 sunshine hours per year. Due to the large geographical spread, the temperature and humidity vary widely in different regions. According to the Bureau of Indian Standards (BIS), the country is classified into five major climatic zones: hot and dry, warm and humid, temperate, cold, and composite. The details of climate zones are below in Table 2.1. It is well established that the energy output of the PV system depends upon the incident solar radiation. It varies throughout the year due to the earth's elliptical orbit with a varying predictable distance between them.

Table 2.1: Climate Zone Classification

Climate (Zones)	**Mean temperature (°C)**	**Relative Humidity (%)**
Hot and Dry	> 30	< 55
Warm and	> 30	> 55
Humid	> 25	> 75
Temperature	25-30	< 75
Cold	< 25	All values
Composite	When 6 months or more do not fall within any of the above category	

The various methods have been suggested by the researchers for the estimation of solar radiation. Andrew A. Lacis et al. suggested that the radiation is absorbed by the atmosphere and the earth's surface. The principal absorbers are water vapour and ozone. The absorption effect of oxygen (O_2), carbon dioxide (CO_2) and aerosol are not considered in this investigation [58]. Thornton, P. E et al. modified the M1 model suggested by Bristow and Campbell in 1984, which identifies a relationship between daily near air surface temperature and daily total solar radiation incidents

on the surface. They suggested including the humidity and precipitation along with the temperature. The result is superior to M1 and covers a wide range [59]. The first theoretical model for estimating global solar radiation based on sunshine duration was given by Angstrom [60]. After that many mathematical linear models presented by the researchers to estimate the solar radiation in [61–67]. A commonly used linear model for solar estimation is given by Equation 2.2.1

$$\frac{H}{H_o} = a + b\left(\frac{S}{S_o}\right) \tag{2.2.1}$$

Where $H, H_o,\ S$ and S_o are global solar energy, extraterrestrial solar energy, day length, and the number of shining hours, respectively; for the better estimation of solar radiation, cloud and atmospheric conditions are considered with a modification in the linear model. A nonlinear term is added to the Angstrom model to obtain the quadratic model given by Equation 2.2.2.

$$\frac{H}{H_o} = a + b\left(\frac{S}{S_o}\right) + c\left(\frac{S}{S_o}\right)^2 \tag{2.2.2}$$

A nonlinear model-based estimation work was reported in [68–72]. Statistical techniques like the Persistence model, ARMA, and ARIMA are also reported in [1, 6, 73–75]. Here, prediction is based upon the requirement of the time series data of meteorological variables, further assuming that the data is generated by a linear process, but it is not true as all practical problems generate nonlinear data. The ability to adapt to nonlinearity and approximate complex relationships without extensive data or knowledge attracts the world to use ANN. An extensive study using ANN for solar estimation was presented in [76–83]. Fuzzy-based prediction models are also reported [36,84–87]. Adaptive Neuro-Fuzzy Inference System (ANFIS) approach, which is a combination of ANN and Fuzzy, also explored for estimating the solar power by researchers [88–91].

Support Vector Machine (SVM) is another popular technique developed by Vipnik [92] used by the researchers. SVM classifies the data with a large margin and thus has better accuracy. It can also be used for performing prediction, using support vector regression (SVR). The performance depends upon the kernel function and the parameters used in SVR. Three kernel functions are mainly used in SVM: Linear

kernel, Polynomial kernel, and Radial Basis Function (RBF) kernel. RBF kernel had better results than the other two because it gives the liberty to play in an infinite space where highly nonlinear data is also separable. The SVR based techniques used for solar radiation prediction are presented in [93–96].

2.3 Support Vector Regression

Support Vector Machines (SVM) primarily solve the problem based on classifications, but they can also perform prediction. The Support Vector Regression (SVR) is the type of SVM capable of predicting. The algorithm for the regression problem is modified by adding the new variable (epsilon) ϵ, which controls the deviation from the desired response. Equation 2.3.1 gives the loss function, which minimizes the error for the prediction.

$$L_\epsilon(d,y) = \begin{cases} |d-y| - \epsilon & for |d-y| \geq \epsilon \\ 0 & otherwise \end{cases} \tag{2.3.1}$$

Here d is the actual output, and y is the estimated value of solar energy based upon the input data set x; here, meteorological parameters. The loss function is equal to zero if the absolute value of the deviation of the assessed value y from the actual value d is less than or equal to zero. The loss function is equal to the absolute deviation value minus ϵ, as depicted in Figure 2.1. Thus for a given observation of N input and output samples, the data set is given $(x_1, d_1), (x_2, d_2) \ldots (x_N, d_N)$, where data is statistically independent and identically distributed.

To solve the nonlinear problem, as in the case of solar prediction, the input vector is mapped in high dimensional feature space and correlated linearly with desired value to form the model given by Equation 2.3.2.

$$y_i = f(x_i) = w^T.\phi(x_i) + b \tag{2.3.2}$$

Here, w is the weighted unit vector normal to the hyperplane, and b is the bias distance from the origin to the hyperplane in high dimensional feature space. The main objective of this problem is to find the optimal weights, and bias along the

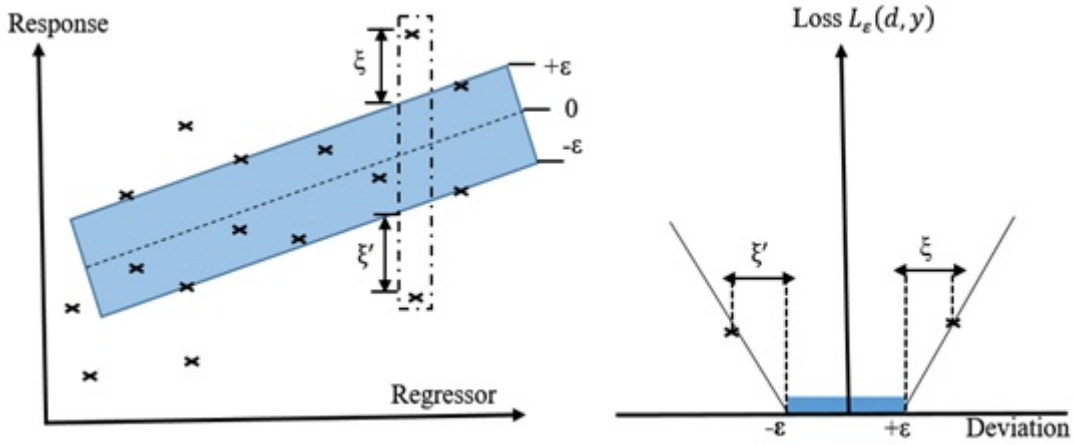


Figure 2.1: Loss function for SVR

error generated in the prediction known as empirical risk should be minimized. The problem can be formulated as given in Equations 2.3.3 and 2.3.4.

$$R_{reg}(f) = min\left\{\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}|y_i - d_i|_\epsilon\right\} \tag{2.3.3}$$

subjected to

$$\begin{cases} d_i\left\{w^T.\phi(x_i)+b\right\} - y_i \leq \epsilon + \xi_i \\ y_i - d_i\left\{w^T.\phi(x_i)+b\right\} - y_i \leq \epsilon + \xi_i^{'} \\ \xi_i\xi^{'} \geq 0 \\ for\ \ i = 1, 2, \ldots, N \end{cases} \tag{2.3.4}$$

The slack variables $\xi, \xi^{'}$ are introduced to cope with the feasibility of the constraints and allow the permissible error, which can be tolerated. The constant C > 0 determines the trade-off between the empirical risk and the regularization capability of the model. To solve for the optimal values of weights and bias lagrangian function is formed given in Equations 2.3.5 and 2.3.6. They are minimized w.r.t. regression parameters w and b along the slack variables $\xi, \xi^{'}$. The parameters γ and $\gamma^{'}$ are added in the function to ensure optimality constraints imposed on the lagrangian multiplier,α and $\alpha^{'}$ which can be calculated through quadratic programming. The values of α and $\alpha^{''}$ can be calculated using the Karush-Kuhn-Tucker condition, which implies that only a few coefficients will be nonzeros for which support vectors exist and lie on the boundary of the optimal hyperplane of regression and optimal values of the unit

vector along with bias is given in Equation 2.3.7 and 2.3.8.

$$minJ(w,\xi,\xi^{'},\alpha,\alpha^{'},\gamma,\gamma^{'}) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi+\xi^{'}) - \sum_{i=1}^{N}(\gamma_i\xi_i - \gamma_i^{'}\xi_i^{'}) - \sum_{i=1}^{N}\alpha_i(w^Tx_i + b - d_i + \epsilon + \xi_i) - \sum_{i=1}^{N}\alpha_i^{'}(d_i - w^Tx_i - b + \epsilon + \xi_i^{'}) \tag{2.3.5}$$

$$Subjected\,to\,\sum_{i=1}^{N}(\alpha_i - \alpha_i^{'}) = 0 \;\; with\, 0 \leq \alpha_i \,and\, \alpha_i^{'} \leq C \tag{2.3.6}$$

$$\hat{w} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^{'})x \tag{2.3.7}$$

$$\begin{aligned} \hat{b} &= d_i = \hat{w}^Tx_i = \epsilon \; for\; 0 < \alpha_i < c \\ \hat{b} &= d_i = \hat{w}^Tx_i = \epsilon \; for\; 0 < \alpha_i^{'} < c \end{aligned} \tag{2.3.8}$$

The prediction done by the regression model can be obtained by substituting all the values in Equation 2.3.3 as given by Equation 2.3.9.

$$R_{reg}f(x) = \sum_{i}^{l}(\alpha_i - \alpha_i^{'})K(x,x_i) + b \tag{2.3.9}$$

The term $K(x,x_i)$, known as kernel function, describes the inner product in the D-dimension feature space given in Equation 2.3.10.

$$K(x,x_i) = \sum_{i=1}^{D}\phi_j(x)\phi_j(d) \tag{2.3.10}$$

2.4 Artificial Neural Network (ANN)

The neural network has been successfully used for estimating the solar energy by many researchers. The ability to deal with non linearity and ability to solve complex relationships deprived of extensive data or knowledge attract the world to use ANN. It eliminate prior requirement for structure of a model which is necessary in

case of statistical methods. Neural network has the capability to perform the various operation like, pattern detection, optimization, clustering and nonlinear estimation. In a traditional NN structure, each input is multiplied by a connection weight. The products are then added together and routed through a nonlinear transfer function to obtain an output. Typically, the result is obtained as the threshold output. The transfer function is a linear or nonlinear algebraic equation. The training step involves updating and determining the connection weights. An epoch is a complete update calculation that includes all of the training patterns. The system refreshes the network parameters at the conclusion of each epoch. Typically, several such epoch passes are required to get the desired optimum. Different connection styles and learning algorithms exist. The Back Propagation technique is a popular and straightforward algorithm that, as expected, includes two phases: training and recall. The network's weights are randomly initialized prior to the training phase. The network's output is then calculated and compared to the target value. The error of the network is calculated at each stage of training and utilized to modify the weights of the output layer using gradient algorithms. If the network has multiple layers of connections, network errors are propagated backward and used to update the weights of the preceding layers. After multiple training steps, the weights are calculated, and the recall phase begins, in which the network output computations are conducted using the input data and weights from the training phase. Once the weight update values are determined, the new weights and biases can be obtained using Equation (2.4.1)

$$W_{ij,n} = U_n + \alpha W_{iJ,n-1} \tag{2.4.1}$$

where $W_{ij,n}$ is a vector of current weights and biases, α is the learning rate which determines how the past weights will reflect to the current value, and U_n is the update function.

2.5 Adaptive Neuro-Fuzzy Inference System

An adaptive network-based fuzzy inference system is a network that facilitates the use of neural network in conjunction with fuzzy logic. ANFIS system is composed of typical fuzzy system components, with the difference that calculations at each

level are conducted by a layer of hidden neurons, and the neural network's learning capacity is provided to build system knowledge. It can handle complex non-linear and dynamic systems. ANFIS is a type of adaptive multi-layer feed-forward network that functions in the same way as a fuzzy inference system. ANFIS can determine the optimal distribution of membership functions by analysing the mapping relationship between the input and output data, either alone or in combination with a least-squares method. An ANFIS network, like ANNs, is trained using supervised learning to go from a specific input to a certain desired output.

2.6 Performance Monitoring Parameters

Various statistical measures are used to evaluate the proposed technique's performance satisfactorily. The performance metrics selected here are Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), and the Coefficient of Determination (R2), as given in equations 2.4.1, 2.4.2, 2.4.3, 2.4.4, and 2.4.5 respectively.

$$MAE = \frac{1}{N}\sum_{1}^{N}|P_{Actual\,Value} - P_{Estimated\,Value}| \tag{2.6.1}$$

$$MAPE = \frac{1}{N}\sum_{1}^{N}|\frac{P_{Actual\,Value} - P_{Estimated\,Value}}{P_{Actual\,Value}}| * 100 \tag{2.6.2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{1}^{N}(P_{Actual\,Value} - P_{Estimated\,Value})^2} \tag{2.6.3}$$

$$NRMSE = \frac{1}{P_{Estimated\,Value}}\sqrt{\frac{1}{N}\sum_{1}^{N}(P_{Actual\,Value} - P_{Estimated\,Value})^2} \tag{2.6.4}$$

$$R^2 = \frac{1 - SS_{Regression}}{SSTotal} \tag{2.6.5}$$

The RMSE represents the variation in the estimated value around the actual value. Another terminology used is a normalized or relative error for comparing the estimat-

ing performance of the differently sized sun-fueled plant by relating the error to the installed capacity preinstalled, measuring NRMSE. The coefficient of determination (R^2)is another measure used to evaluate the model's performance. The value of R^2 lies between 0 and 1. The value closer to 1 reflects that model performance is very good, and the prediction was accurate.

2.7 Methodology for Estimation

The complete procedure for estimating solar power using SVR and its refinement by evaluating the degradation in solar panels with the help of the clustering-based technique is shown in Figure 2.2. The clustering and degradation part will be discussed in chapter 3. Here, we will cover the estimation part shown at the triangle's base in Figure 2.2. The steps used for the solar power estimation is shown via a flowchart diagram in Figure 2.3.

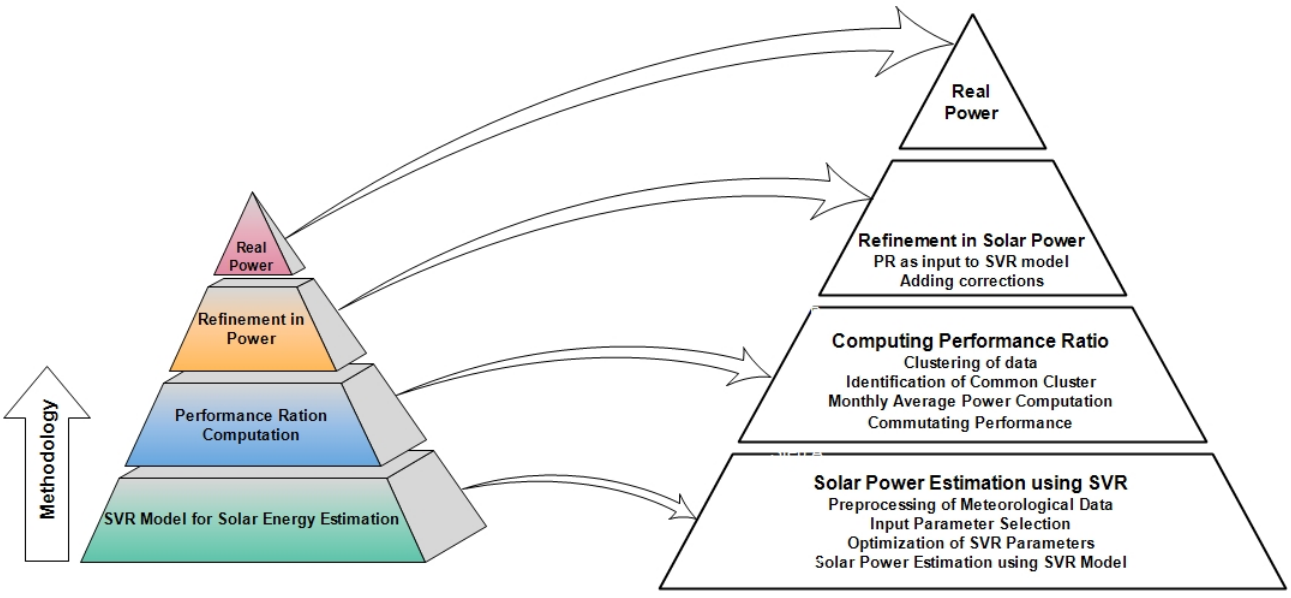


Figure 2.2: Methodology for Real-Time Estimation of Solar Power

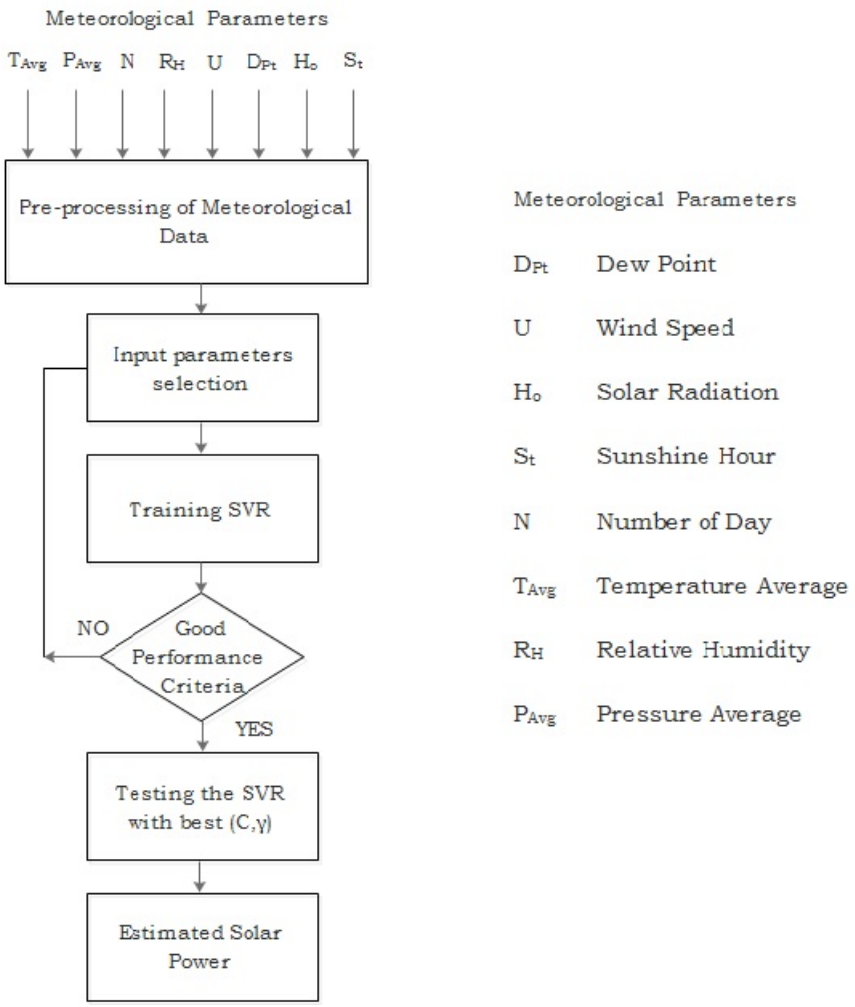


Figure 2.3: Solar Energy Estimation using SVR

2.7.1 Pre-processing of Meteorological Parameters

This section refers to the data collection, preparation, and cleaning part. The section is discussed in detail in chapter 4. However, some of the important steps are mentioned here also.

Interpolation for the missing values in the meteorological data:

The input dataset obtained from the plant is corrupted; some values were duplicated, and some were missing. To use this data set, duplicate values were removed, and interpolation was applied to estimate the missing values.

Conversion of data from minute to daily format: The per-minute meteorological data are obtained from the plant, but as per the scope of the proposed model, the data are converted hourly and finally into the daily format by a simple averaging technique.

Normalization of data: The input meteorological data have large variations in their parameter values, which is difficult to model. Therefore, to limit the data to a uniform range [0 1], we normalized it using the max-min normalization (also known as feature scaling).

2.7.2 Input Parameter Selection

The correlation coefficient is used to find the most suitable parameters for solar energy estimation. We have checked the dependency of different meteorological parameters on solar radiation/power. For this, we have used Pearson's correlation coefficient. The correlation coefficient (r) value lies between -1 and 1. The parameters showing the positive value of r and closer to unity show a linear relationship between input parameter with solar radiation/Power shown in Table 2.2.

Table 2.2: Pearson's Correlation Coefficient Matrix

Correlation Coefficient	**Radiation**	**Temp.**	**Humidity**	**Wind Speed**	**Pressure**	**Dew Point**
Radiation	1.00	0.66	-0.47	0.58	0.25	-0.08
Temperature	0.66	1.00	-0.95	0.82	-0.42	-0.64
Humidity	-0.47	-0.95	1.00	-0.80	0.65	0.83
Wind Speed	0.58	0.82	-0.80	1.00	-0.39	-0.60
Pressure	0.25	-0.42	0.65	-0.39	1.00	-0.39
Dew Point	-0.08	-0.64	0.83	-0.60	-0.39	1.00

2.7.3 Optimization of Kernel Parameter

we have used the Radial Basis Function (RBF) kernel for estimating solar power. Two essential parameters of kernel function (C and γ) require the optimized value for estimation. The trading mistake penalty for stability is determined by parameter C, which regulates each support vector's influence. Parameter γ compensates between error due to bias and variance. The grid search technique is used for obtaining optimized values of C and γ. Table 3 reports the optimized values of C and γ for the best models out of 52 models.

Table 2.3: Optimized Values for the Best Fit Models

Input combination	C	γ
$N, T_{avg} R_H, H_0, S_t, PR, (SVR_{35})$	0.5	8
$N, T_{avg}, H_0, S_t, PR, (SVR_{36})$	128	0.125
$N, T_{avg}, P_{avg}, H_0, S_t, PR, (SVR_{37})$	128	0.125
$N, H_0, S_t, PR, (SVR_{48})$	0.5	2

2.7.4 Solar Power Estimation using SVR

The complete process of modeling for solar power estimation using support vector regression is shown in Figure 2.3. The meteorological parameters act as the input to the SVR model, while solar power is the output. After pre-processing, and feature selection, different models are studied. Once the best models are known, they are optimized with the best values of C and γ via the grid search method. Finally, the model performance is evaluated using the test data.

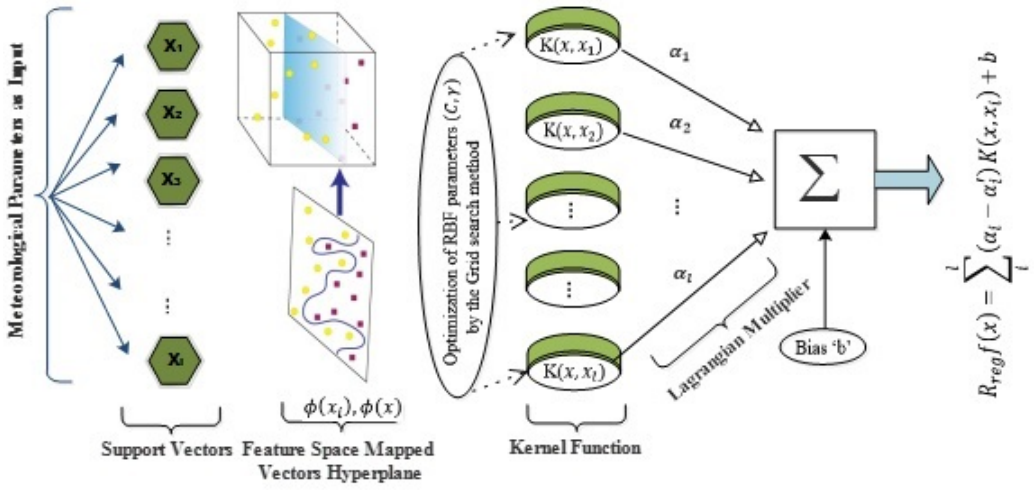


Figure 2.4: Support Vector Regression model for Solar Energy Estimation

2.8 Summary

After going to the various methods reported in the literature for solar energy estimation, the SVR technique is applied for estimating solar irradiance and solar power. The LIBSVM library for SVM is used here in the MATLAB environment. The meteorological data collected from the NISE Gurgaon is converted into LIBSVM format to predict the solar Irradiance and solar power. The different input combination me-

teorological parameters are applied to find the best input combination. The SVR parameters are also optimized via the grid search method to enhance prediction accuracy further.

Chapter 3

Clustering based Computation of Degradation Rate (CCDR)

The chapter is dedicated to degradation induced in PV panels. Here, the different methods of computing the degradation in PV panels are discussed. A new method to find the degradation induced in panels is presented here.

3.1 Introduction

The world is currently engulfed in a frenzy to install photovoltaic (PV) systems in deserts, over water bodies, on rooftops of houses, automobiles, and parking spots, among other places [97]. Due to continued investment and innovation in technology, project funding, and execution, global PV installation is anticipated to have crossed 600 GW to date and is expected to cross 4500 GW by 2050. Therefore, performance monitoring of the PV system is an excellent area of interest to work for the researcher's community [98–101]. PV technology's performance is reflected in degradation, a gradual decrease in power output from different PV panels over the years. Due to the degradation in PV panels, it becomes tough for researchers to forecast solar energy accurately. For accurate energy forecasting, the degradation induced in solar panels should be considered. Moving in this direction, we have developed a heuristic technique to calculate the degradation induced in solar panels in a time without physically going on site, which is further used to fine-tune the solar power output.

This chapter describes a new machine learning-based method to calculate degradation: Clustering-based Computation of Degradation Rate (CCDR).

3.2 Literature Survey of Degradation

3.2.1 Background

The degradation in the performance of PV panels can occur at the cell, module, or array level. The main factors responsible for degradation at the cell level include temperature, humidity, precipitation, snow, dust, and solar radiation. Delamination, corrosion, discoloration, module mismatch, and shading are the major sources of degradation at the array level. Figure 3.1 contributes to different PV modules degradation modes. The major contribution occurred due to the delamination (42%)

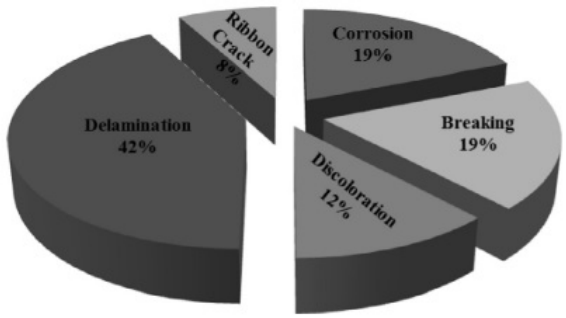


Figure 3.1: Different Modes of Degradation in PV Panels

depicted in Figure 3.2. According to [102], delamination is the lack of adhesion between the cells' encasing polymer or between the cells and the front glass. This poses a significant issue since it increases light reflection and allows water to seep into the module's construction. The edges of the module are where delamination is most severe because, in addition to causing power degradation, it also puts the module and the installation at risk for electrical failure [103]. In hot and humid environments, delamination frequently occurs [104]. The location considered in our study also lies in a hot and humid environment. Discoloration changes the material's colour in the module to yellow and sometimes to brown colours as depicted in Figure 3.2. The power produced by the module is decreased due to the change in how much light is transmitted to the PV cells [105–107]. Another significant contributor to PV module degradation is glass breakage [108]. It often happens during installation, servicing, and particularly when transporting modules to their installation sites reflected in Figure 3.3. Even if the equipment is damaged, it can still function normally. A PV module with cracks

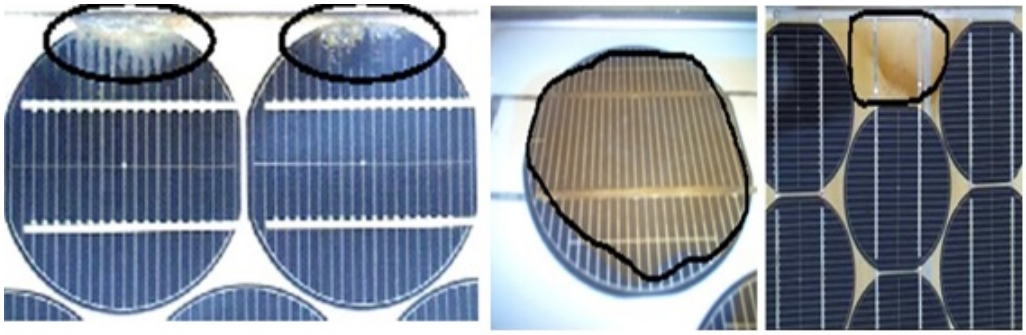


Figure 3.2: Delamination and Discoloration in PV panels

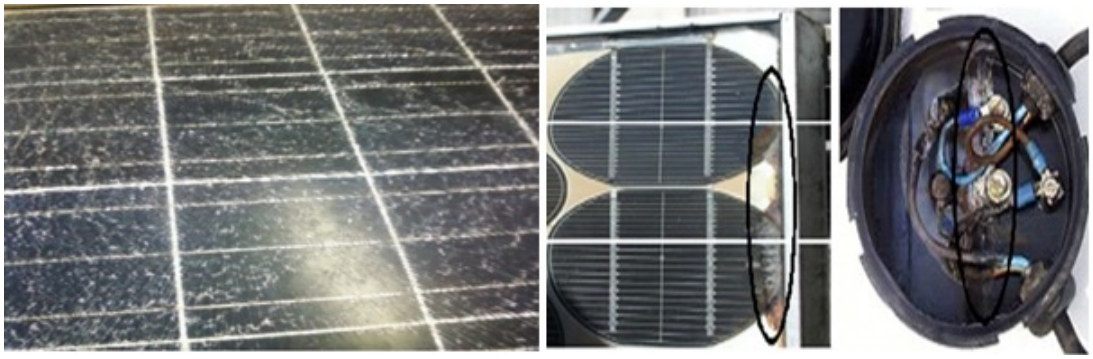


Figure 3.3: Broken Glass and Corrosion in Panels & Box

can operate for many years without experiencing a substantial loss of energy. However, it raises the threat of electric shock and moisture interference [109]. Breakages and cracks are typically followed by other deterioration like corrosion, discolouration, and delamination. The moisture also moves inside the module through breaking glass, causing corrosion [110]. The metallic connectors of PV cells are attacked by corrosion, causing an increase in leakage currents [111]. Additionally, corrosion weakens the bond between cells and metallic frames [112]. A PV module with corrosion at the edge and junction box is depicted in Figure 3.3. An experimental study collected data from 57 crystalline silicon modules in five different climatic zones of India. The study concluded that discoloration-induced degradation is more prominent in hot and dry climate zones, while corrosion is more important in hot and humid zones [113, 114]. It was also concluded that cold climatic zones show the least degradation. Numerous experiments have been conducted to explain degrading and failure patterns, resulting in new approaches to look into the different degradation types. These studies helped to improve the lifetime of PV modules. To inspect the failure and degrading modes of PV panels, which have been working for the last 10 to 15 years in the real environment, the following are standard methods like visual inspection and imaging

techniques.

3.2.1.1 Visual Inspection

The visual inspection approach is used to find degradation modes or faults that are visible to the naked eye in older PV panels. This method can identify some PV module degradations, such as encapsulation discolouration, delamination, glass breaking, bubbles, corrosion, bus bar oxidation, frame faults, etc. [114, 115].

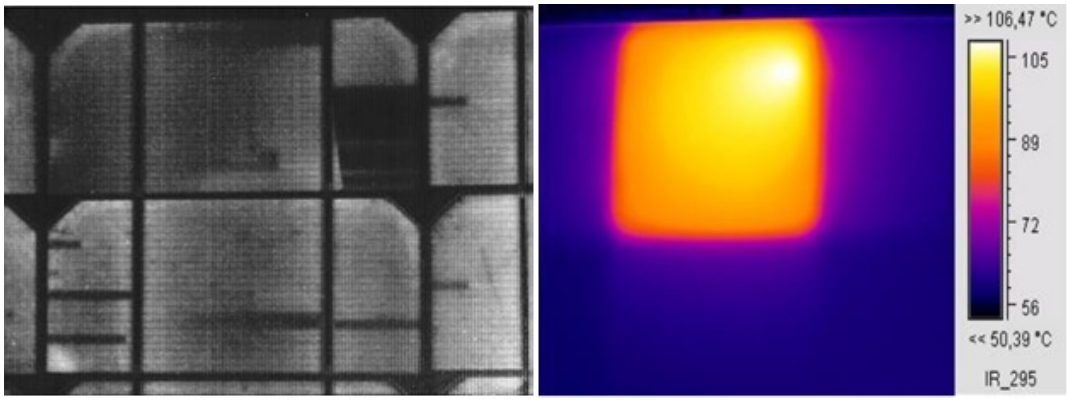


Figure 3.4: EL Image Showing Cracks & IR Image Showing Hot Spot in Panels

3.2.1.2 Imaging

Electroluminescence imaging (EL) is the most frequently used technique to identify cell material cracks, such as micro-cracks, fractured metallization, shunts, and inactive regions. It helps distinguish between increasing series resistance and lower shunt resistance [116]. A micro crack in a cell and numerous broken fingers are seen in the EL image of a PV module showing early signs of aging after several hundred operating hours, which was captured using a SWIR in GaAs sensor with a spectral range of 0.9 to 1.7μm (Figure 3.4). Thermal and infrared imaging is another method to locate the degradation or failure spots that are not visible in the old-aged module. E. Kaplani [117] highlighted the corrosion on the bus bar via IR thermography as an intense hot spot, as depicted in Figure 3.4. The temperature of the hot spot region is 40°C, higher than the module's average temperature. IR images were captured with the help IR camera having a spectral range of 7.5–14μm. Hot cells operating at 20–30°C show significant temperature changes that indicate the cells' defect, as shown in Figure 3.4.

3.2.2 Performance monitoring of PV panels

As the installation cost of PV systems is comparatively higher than the conventional systems, the estimate of energy output at the site and the performance monitoring of PV panels are required to utilize solar energy effectively. The parameters used to reflect the performance of PV panels are depicted in Figure 3.5. The Capacity Utiliza-

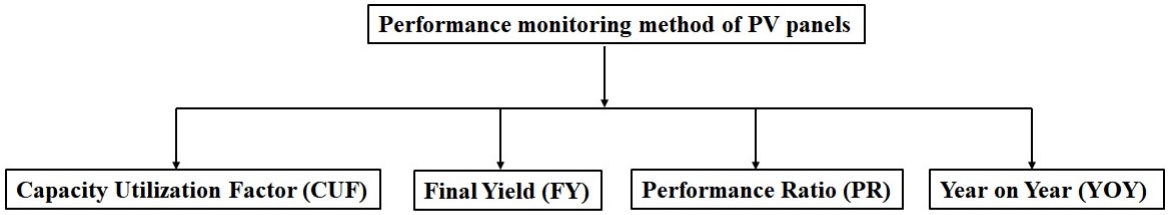


Figure 3.5: Performance Monitoring Methods for PV Panels

tion Factor (CUF), Final Yield (Y_F), Year on Year (YOY), and Performance Ratio (PR) are the most common parameters to evaluate the performance of PV systems by industries. The capacity utilization factor represents the capacity of a PV plant in a certain period without considering environmental factors. If the system constantly produces power at its rated capacity, the CUF will be unity. Equation 3.2.1 provides the formula for it, which is calculated as the PV system's actual annual energy production (E_{ac}) divided by the amount of energy the PV system would produce if it were maintained at full rated power P_m for 24 hours a day for a year.

$$CUF = \frac{Y_F}{365 * 24} = \frac{E_{ac}}{P_m * 8760} \tag{3.2.1}$$

The final yield of the system Y_F is determined by dividing the system's final to actual energy output by its nominal D.C. power given in Equation 3.2.2 The Y_F scales the energy output according to the system size. The final yield represents the number of hours per day the array would need to operate at its rated output power Pm to equal its monitored contribution to the net daily load.

$$FinalYield(Y_F) = \frac{Final\ Energy\ output\ KWh}{Nominal\ DC\ Power\ KW} \tag{3.2.2}$$

The ratio of the PV's total in-plane irradiance (H) to the reference irradiance (G) given in Equation 3.2.3 is known as the reference yield (Y_R). In other words, it is equal to the quantity of hours at the reference irradiance to the number of hours during the lenght of the summer sun. It serves as a representation of the solar resource available to the PV system at that location. It is influenced by the PV array's orientation and seasonal and monthly weather variations.

$$ReferenceYield\ (Y_R) = \frac{H}{G} \tag{3.2.3}$$

Performance Ratio (PR) is advantageous over other commonly used techniques as it indicates the actual energy delivered by the plant instead of theoretical value under a given insolation and climate condition. PR indicates losses resulting from the cell/array mismatch, shading, inverter problems, module temperature, etc. In addition to being used to compare grid-connected PV systems regardless of their location, power capacity, and mounting structure, it is the most often utilized as a performance metric for comparative analysis of various PV panel technologies. It is thus given by the ratio of the final yield to the reference yield given in Equation 3.3.4. The value of PR treats good with a figure of around 0.8–0.85. The values less the 0.75 indicates the poor performance of the PV system, and it needs to be corrected.

$$PerformanceRatio(PR) = \frac{Y_F}{Y_R} \tag{3.2.4}$$

Similarly, the PR has been calculated using the Robust Principal Component Analysis (RPCA) technique [118]. Three technologies were examined in that study: monocrystalline silicon, multicrystalline silicon, and hetero-junction with an intrinsic thin layer (HIT). When monthly PR was determined for eight years of plant data collected at various places throughout Cyprus, the results for thin-film technology were reassuringly accurate. The degradation of rooftop PV systems installed at the National Renewable Energy Laboratory (NREL), USA, was examined after 20 years of operation in [119]. The degradation rate of two mono-Si arrays was reported at 0.8 percent annually using the mean average PR method. In a similar finding, the PR degradation of 90 mono-crystalline PV modules that had been installed for 22 years on the roof of the NISE guesthouse in Gurgaon, India, was determined [120]. It was dis-

covered that the average deterioration rate in this example was 1.9 percent per year, with a maximum reported value of 4.1 percent per year and a minimum reported value of 0.3 percent per year. Another investigation of thermal damage was done elsewhere, utilizing effective temperature and meteorological information [121]. Reference [122] demonstrated the deterioration impact of PV modules that are opaque and semi-transparent. Semi-transparent modules have a lower deterioration rate than opaque modules. The degradation of PV panels can also be computed with the help of regression and the year-to-year (YOY) method using PR. Linear Regression (LR) is the most common method for calculating the degradation, which aims to minimize the sum of squared residuals reported in [123]. However, this method is susceptible to outliers and seasonal variation and thus has enormous uncertainty. The Classical Seasonal Decomposition (CSD) method is used to overcome this limitation and extract the trends from PV time series data [124, 125]. Using centered moving average computation and the presumption that seasonal components are stable across time, this approach extracts the seasonal component of each month. Due to large autocorrelations in the model residuals, which occur when the LR and CSD approaches are fitted to a fixed model, they cannot capture some of the key characteristics of solar energy time series. Another statistical decomposition method based on locally weighted scatterplot smoothing (LOESS) is proposed in [126]. The Solar Energy Institute of Singapore evaluated the performance of nine different PV panels. Monocrystalline and amorphous silicon technologies were proven to perform better than multi-crystalline technology. The deterioration rate for the copper-indium-gallium-selenide (CIGS) module was the highest at 6% annually. Compared to CSD or ARIMA, LOESS offers more reliable estimates for seasonal components and trends. From the above analysis, it was clear that the regression method requires filtering, as it is sensitive to outliers; therefore, the Year-On-Year (YOY) method is preferred to overcome the filtering requirement. The year-on-year (YOY) comparisons are a popular and helpful technique for evaluating a solar PV panel's performance. On a year-over-year basis, any repeating quantifiable event can be compared. Performance on a yearly, quarterly, and monthly basis are common YOY comparisons. The YOY method is also insensitive to outliers but requires multiple years of data presented in [127, 128].

3.3 Methodology for CCDR

The proposed work suggests a new approach for computing PV module degradation employing the unsupervised clustering technique. There is no requirement for a physical inspection of the module on the site, and the procedure only requires knowledge of historical weather data. Degradation rate is a significant measure representing the plant's performance and is defined as the decrease in power output for the same input conditions over time. The current technology uses clustering's strong pattern recognition capabilities to obtain similar input conditions. The method uses three years of meteorological data (2010 to 2012). We have used the K-means clustering algorithm for this work.

3.3.1 K-mean Clustering

The main task of clustering is to group the given data set into groups or clusters so that the data within a cluster are similar and the data inside other clusters differ. One such clustering technique has drawn interest from the scientific community is the k-means clustering technique. It helps in identifying any given data. This method of learning is unsupervised. K-means has several uses in several fields of study, particularly in predicting renewable energy. One can better understand the data and increase estimation quality by managing sets of data individually. The k-means clustering technique is widely used for the clustering of data. The formation of clusters is done using the centroid technique, which arranges the data in a cluster for a given centroid. Finally, the Error function is minimized as per Equation 3.3.1

$$E = \sum_{j=1}^{N} \sum_{p \in C_i} dist(p, C_i)^2 \tag{3.3.1}$$

Where E error function for the entire object in the dataset, p represents all objects point in space, C_i is the cluster's centroid, and K is the Number of clusters. Let X_i having N observation represent a collection of multidimensional meteorological data that will be segregated into a set of K clusters, where K is less than the observation N. The squared Euclidean norm is used to define the similarity measure between each pair of input data X_i and $X_{i'}$ vectors in K-means clustering given by Equation

3.3.2. The Euclidean distance is the straight line distance between two points under consideration. The Manhattan distance is another method for finding the clusters. It is the city block distance between the two points. However, it is not a shorter distance. The Manhattan distance is given by Equation 3.3.3.

$$d(X_i, X_{i'}) = |X_i - X_{i'}|^2 \tag{3.3.2}$$

$$d(X_i, X_{i'}) = |X_i - X_{i'}| \tag{3.3.3}$$

The meteorological parameters are used as the input vector. K-means clustering algorithm determines similar input conditions in data from 2010-2012. The clustering algorithm assigns a cluster number to each value of input data. The specific cluster of meteorological data provides the same uniform input conditions for all three years, i.e., 2010–2012, as shown in Figure 3.6. The complete procedure for the CCDR tech-

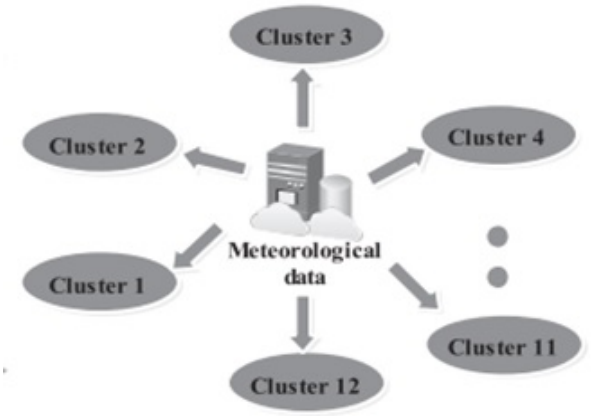


Figure 3.6: Clustering of Meteorological Data

nique is shown in Figure 3.7. The proposed technique uses pre-processed data for the degradation rate calculation. Finding patterns or clusters of similar meteorological conditions in the input data occur after pre-processing. For all three years, from 2010 to 2012, and for all the PV topologies under examination, the weather data in a particular cluster give roughly uniform input circumstances. The current work used the traditional K-means clustering method to find these clusters in the data. K's value is carefully selected to account for seasonal changes throughout the year. Summer, the wet season, autumn, winter, and spring are the four main seasons where the plant is

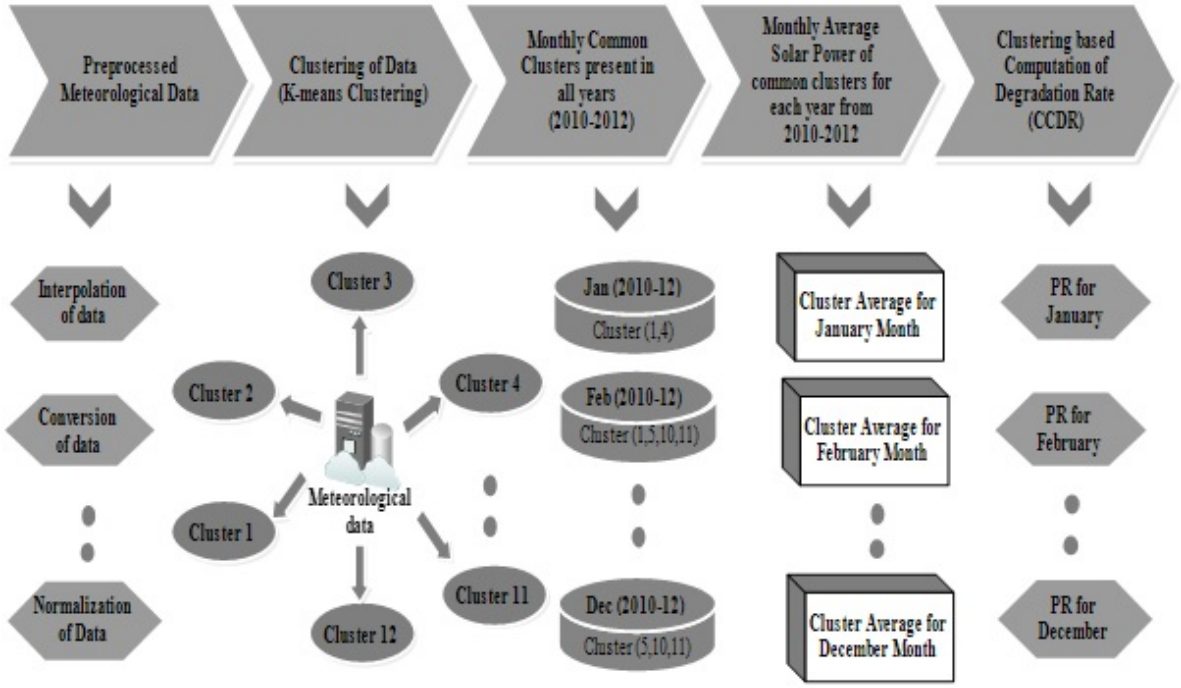


Figure 3.7: Clustering-based Computation of Degradation Rate

located. Experiments were conducted in which a range of values of K was considered. The optimized value of K was found to be 12, which covered all the seasonal variability at the plant. It is important to mention that the clustering algorithm was applied to the whole data set for 2010-2012. However, similar input patterns for each month are required to calculate the degradation rate. Therefore, considering the seasonal variability, we have further subdivided the clusters according to their month for all three years. This step arranges the whole data set monthly, with each monthly group (also called the common cluster) showing the same weather conditions. By independently averaging the common clusters for each year after knowing the monthly data groups, it is possible to calculate the equivalent solar power. This procedure provides the monthly average solar power for similar input conditions each year. This concept is used to find the change in power, thus, the degradation rate. The suggested methodology is illustrated for a small portion of the data to make it easier to understand. The following steps provide a summary of the entire methodology:

3.3.1.1 Step-1: Clustering of meteorological data

The input vector includes the outside temperature, humidity, wind speed, pressure, dew point, irradiance, and sunshine hours. The segmental K-means clustering algo-

rithm identifies common patterns in the input data. After implementing the clustering method, the data are allocated with clusters, as depicted in Table 3.1.

Table 3.1: Assigning the Cluster to Meteorological Data

Time	Absolute Temperature	Relative Humidity	Wind Speed	Atmospheric Pressure	Dew Point	Irradiance Horizontal	Sun Shine	Clusters
01/01/2010	12.63	69.81	0.91	986.10	6.44	137.69	7	5
02/01/2010	9.50	91.15	0.94	988.76	8.08	50.25	3	10
03/01/2010	10.20	79.9	0.55	985.10	6.47	126.14	7	6
01/01/2011	10.77	77.33	1.57	984.43	6.73	120.31	6	10
02/01/2011	9.92	70.36	1.78	984.54	4.68	109.68	6	7
05/01/2011	10.20	79.9	0.55	985.10	6.47	126.14	7	6

3.3.1.2 Step-2: Month-wise arrangement of cluster data

Once the cluster number of the meteorological data is assigned, the next step is to arrange the meteorological data according to their cluster number in each year from 2010-2012. Only those clusters are considered in this process, which is common every month from 2010-2012. This helps us to find the condition of similar input conditions required from 2010-2012 shown in Table 3.2.

Table 3.2: Monthly Arrangement of Common Clusters

Time	Absolute Temperature	Relative Humidity	Wind Speed	Atmospheric Pressure	Dew Point	Irradiance Horizontal	Sun Shine	Clusters
01/01/2010	12.63	69.81	0.91	986.10	6.44	137.69	7	10
02/01/2010	9.50	91.15	0.94	988.76	8.08	50.25	3	10
03/01/2010	10.20	79.9	0.55	985.10	6.47	126.14	7	10
01/01/2011	10.77	77.33	1.57	984.43	6.73	120.31	6	10
02/01/2011	9.92	70.36	1.78	984.54	4.68	109.68	6	10
05/01/2011	10.20	79.9	0.55	985.10	6.47	126.14	7	10

3.3.1.3 Step-3: Monthly average solar power

The next step after month-by-month allocation is to process the power outputs associated with the input meteorological vectors. The year is used to categorize the monthly cluster data in this instance. The output power averaged over each month for each common cluster is then determined. This process helps us to determine the monthly average power for each year. The same process is repeated each year from 2010-2012. Table 3.3 displays the monthly average solar power of the a-Si panels.

Table 3.3: Monthly Average Solar Power of a-Si Panels

Month	2010	2011	2012
January	119.65	113.72	102.18
February	193.82	190.39	176.50
March	307.41	276.92	255.36
April	273.73	7251.51	241.36
May	266.86	243.70	236.00
June	265.54	261.97	236.23
July	240.29	229.52	194.68
August	156.98	152.72	145.49
September	192.05	189.51	173.01
October	251.31	242.88	235.76
November	201.74	189.56	182.35
December	166.68	149.46	144.65

3.3.1.4 Step-4: Performance ration calculation

The PR is then determined using the monthly power output for each of the corresponding years determined in step 4 using Equation 3.3.3. The resulting monthly PRs are shown in Table 3.4.

$$PR = \frac{Monthly\ power\ in\ 2011}{Monthly\ power\ in\ 2010} \tag{3.3.4}$$

3.3.1.5 Step-5: Clustering based Computation of Degradation Rate

Finally, the degradation rates for the three separate technology modules are estimated using the PRs determined in step 4. The conventional least-square regression method is used to calculate the deterioration rates. The degradation rate is based on the slope of the regression line for the monthly PR. The degradation rate calculated via the CCDR method is used for fine-tuning the estimation results obtained via the SVR method.

3.4 Summary

In Summary, the methods presented in the literature survey compute the degradation in PV panels either by the physical inspection of the site or via imaging technique. The regression method suffers from outliers, while the YOY method requires multiple years of data. Therefore, real-time information on degradation induced in PV panels

Table 3.4: Monthly Performance Ratio by CCDR

Month	**PR(11/10)**	**PR(12/11)**	**PR(12/10)**
January	0.95	0.89	0.85
February	0.98	0.92	0.91
March	0.90	0.92	0.83
April	0.91	0.95	0.88
May	0.91	0.96	0.88
June	0.98	0.90	0.88
July	0.95	0.84	0.81
August	0.97	0.95	0.92
September	0.98	0.91	0.90
October	0.96	0.97	0.93
November	0.93	0.96	0.90
December	0.89	0.96	0.86

is not possible. The current method has suggested a clustering-based model calculate the rate at which solar panels deteriorate. The main benefit of the suggested approach is that it can calculate the performance ratio of PV panels without requiring on-site physical inspection of the panels. It can therefore be used to estimate degradation in real-time. With the aid of the model, the performance degradation for three PV technologies—polycrystalline silicon, amorphous silicon, and hetero-junction with intrinsic thin-layer silicon—was evaluated in this research work. The findings are comparable with those obtained using other techniques. Additionally, it can be said that the proposed model is quicker and less complex than earlier ones.

Chapter 4

Data and Site Description

The site description, data, and experimental results of solar energy estimation using our proposed methods are presented in this chapter. We have calculated the solar power for the PV panels installed at the NISE Gurgaon. To further increase the estimation accuracy, the degradation brought about in the PV panels was calculated using the CCDR approach covered in chapter 3.

4.1 Site Description

The National Institute of Solar Energy (NISE), a premier institution in the solar energy sector located in Gurgaon, was the site considered for this study. Figure 4.1 show its geographic location at 28° 37'N and 77° 04'E indicates that it is close to the Nation's capital New Delhi. It has a "composite" climate, according to the Bureau of Indian Standards, with dry summers, wet rainy seasons, and chilly winters. The range of the daily average temperature in the summer, from April to June, is 46 to 33 degrees Celsius, with a mean of 41 degrees Celsius. During the rainy season, from July to August, the relative humidity ranges from 42 to 70 percent, which is high for India's annual average. With a yearly average wind speed of 1.5 m/s, the wind is light and moderate. From 2.6 KWP/m^2/day in January to 6.1 KWP/m^2/day in May, the monthly average for horizontal solar radiation changes significantly. Figure 4.2 depicts the experimental setup, which consisted of PV panels using three different technologies: amorphous silicon (a-Si), polycrystalline silicon (p-Si), and heterojunction with thin intrinsic layer (HIT). The HIT array has a maximum rating of 1.68 KWp, the a-Si array of 1.2 KWp, and the p-Si array of 1.6 KWp. According to Figure 4.3, the a-Si PV array has 16 modules of 75 Wp each, the HIT array has 8 modules of 210 Wp each, and the p-Si array has 10 modules with a rated value of 160 Wp.

Figure 4.1: Site Location of NISE Gurgaon

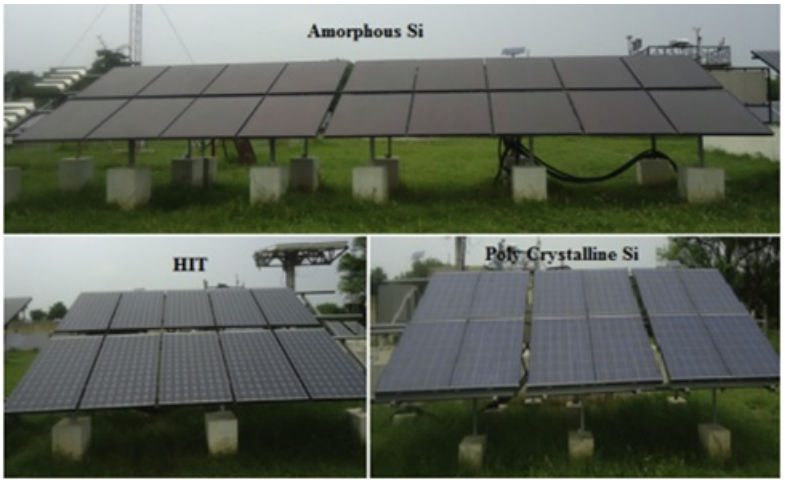


Figure 4.2: Experimental Setup at NISE Gurgaon

4.2 Data Description

For the current research project, we used solar power and meteorological data from 2010 to 2012 by the National Institute of Solar Energy (NISE), India. Every 10 minutes, each PV array's I-V performance data are collected. The analyzer finds the maximum power (Pmax) saved in the data logger. As a result, every 10 minutes, the solar power data for the three technologies are maintained. Most of the weather monitoring equipment used at NISE Gurgaon was made in Japan by M/S EKO Instruments. Every minute of data is recorded for three years by the Campbell Scientific CR-1000 data recorder (2010-2012). The hourly, daily, monthly, and yearly databases for solar power are created by further processing these data. The

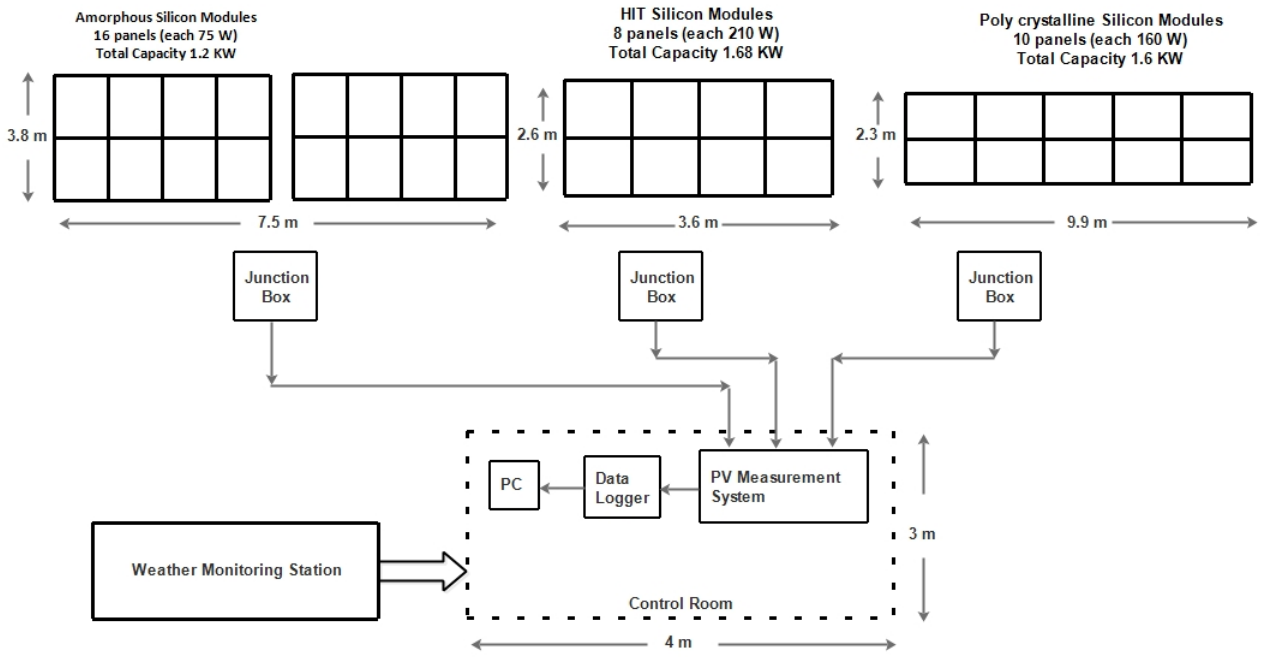


Figure 4.3: Schematic of PV Panels at NISE Gurgaon

meteorological information was acquired from the NISE, Gurgaon weather monitoring station. The variables that the weather station records are the ambient temperature,

Figure 4.4: Weather Monitoring Station at NISE Gurgaon

relative humidity, atmospheric pressure, wind speed, dew point, wind direction, and solar radiation. The database does not contain sunshine data. We computed it using the information on solar radiation. The hourly, daily, monthly, and yearly databases were created by processing the data records for every minute. Modeling is difficult since the data differ significantly across the values of the different parameters. Ta-

ble 4.1 shows the vast difference between the data's minimum and maximum values. Along with data on solar output, figure 4.5 from 2010 to 2012 also depicts the variation of meteorological conditions. We receive 920 observations after pre-processing the data, which involves deleting any redundant or fuzzy values. These values are further separated into 620 training samples, 150 validation samples, and 150 testing samples.

Table 4.1: Range of different Meteorological Parameters

Parameters Range (units)	Absolute Temperature (°C)	Relative Humidity (%)	Wind speed (m/s)	Atmospheric pressure (mmHg)	Dew point (°C)	Irradiance Horizontal (wh/m^2)	Sun Shine (hours)
Minimum	6.17	11.99	0.10	963.51	-4.06	31.84	0.00
Maximum	42.53	93.03	5.00	991.73	25.07	309.19	11.00

4.2.1 Data Pre-processing

The data collected from NISE Gurgaon is raw data. It should be corrected before it can be used in modeling. The daily database is created by processing the minute-by-minute database. We removed the days' worth of meteorological data from the database that had corrupted or missing hourly values. The prepossessing includes the following steps:

Removal of duplicate and Interpolation for the missing values: A few of the values in the input dataset that was collected are missing because it was corrupted. The missing numbers are therefore replaced using the interpolation approach. Each missing value is replaced by a new data point, and the corrupted data points are eliminated from the dataset. The missing values are determined by finding the new data points that correspond to the missing values using linear interpolation. Equation 4.2.1 provides the formula for the interpolation.

$$y - y_1 = \frac{(y_2 - y_1)}{(x_2 - x_1)} * (x - x_1) \tag{4.2.1}$$

*Conversion of data:*The plant provides the meteorological data on a minute-by-minute basis, while the solar power data is obtained after every 10 minutes. However, the data needed to be translated into the hourly format, then into the daily format, which we have done using the averaging technique, to meet the requirements of the suggested model.

*Normalization of data:*It is difficult to model the different ranges of the input vectors

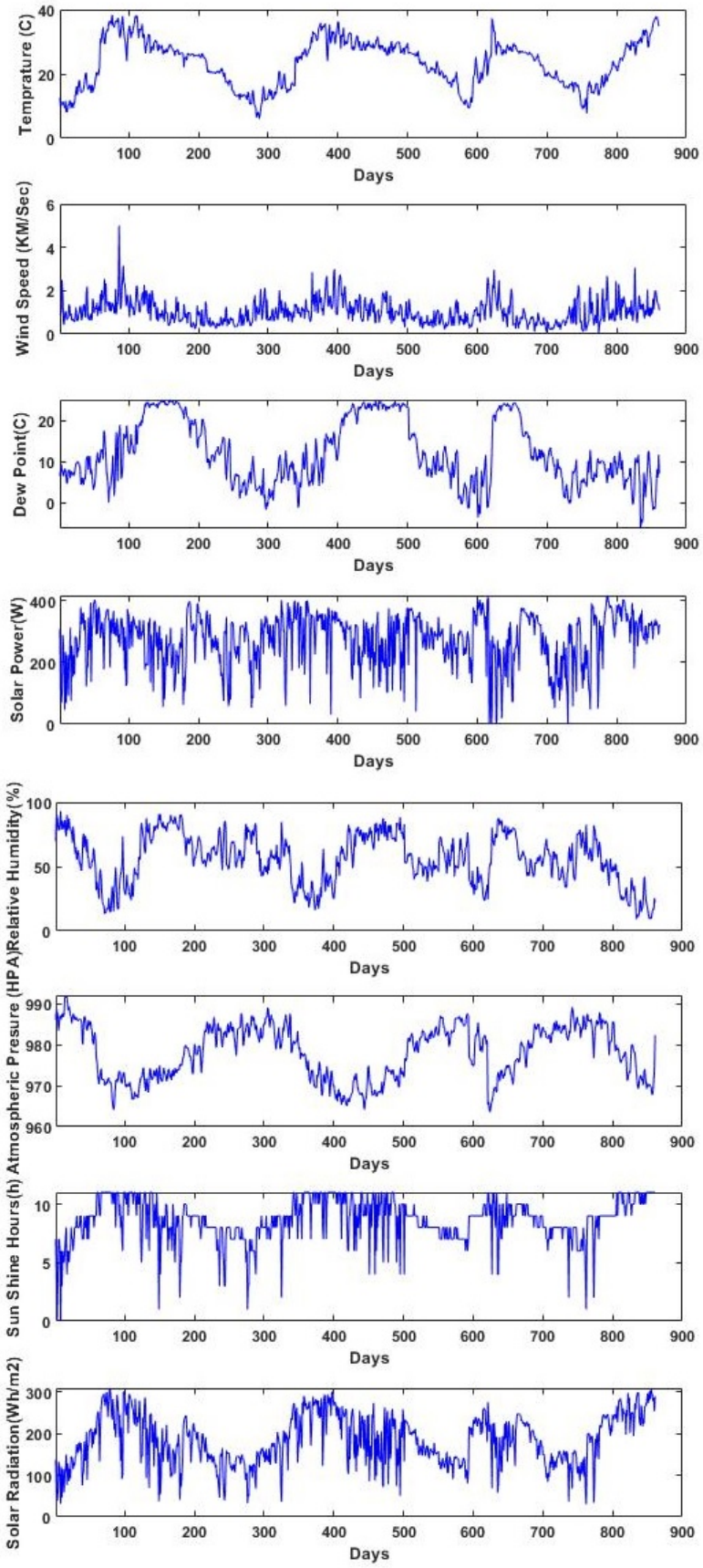


Figure 4.5: Variation of Meteorological Parameters

of the meteorological data. Thus, to limit the data to the same range, we next normalize the data by restricting the input parameters to the range [0 1] using the max-min normalization (also known as feature scaling). Equation 4.2 is used to scale the features. Here, X is the actual value of the meteorological parameters, and X' is the normalized value of the data. In equation 4.2.2, the relation that was employed for feature scaling is described.

$$X^{'} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4.2.2}$$

Where $X^{'}$ is the normalized value of the data, X represents the Input meteorological data, and X_{min}, X_{max} are the minimum and maximum values of the data.

*Feature selection:*The best parameters for solar energy estimation are determined using the correlation coefficient. The correlation between solar power and meteorological variables is shown in Figure 4.6. We examined the relationship between several climatic conditions and solar radiation and power. We have employed Pearson's correlation coefficient for this. The Pearson correlation coefficient (r) ranges from -1 to 1. A linear relationship between the input parameter and solar radiation/power is shown by the parameters with positive r values and values closer to unity. Table 4.2 presents the matrix of Pearson's correlation coefficient along all parameters. Although the suggested model is applied to three years' worth of data, we are simply providing the results for 50 values in Table 2 to make the demonstration easier. It has been noted that the key factors in estimating solar energy are solar radiation, temperature, and sunshine hours. The least appropriate parameters include the dew point, atmospheric pressure, humidity, and wind speed. Different input parameter combinations are utilized for modeling based on the information.

Table 4.2: Feature Selection by Correlation Coefficient Matrix

Correlation coefficient	**Solar Radiation**	**Temperature**	**Humidity**	**Wind Speed**	**Pressure**	**Dew Point**
Solar Radiation	1	0.66	-0.47	0.58	0.25	-0.08
Temperature	0.66	1	-0.95	0.82	-0.42	-0.64
Humidity	-0.47	-0.95	1	-0.80	0.65	0.83
Wind Speed	0.58	0.82	-0.80	1	-0.39	-0.60
Pressure	0.25	-0.42	0.65	-0.39	1	-0.39
Dew Point	-0.08	-0.64	0.83	-0.60	-0.39	1

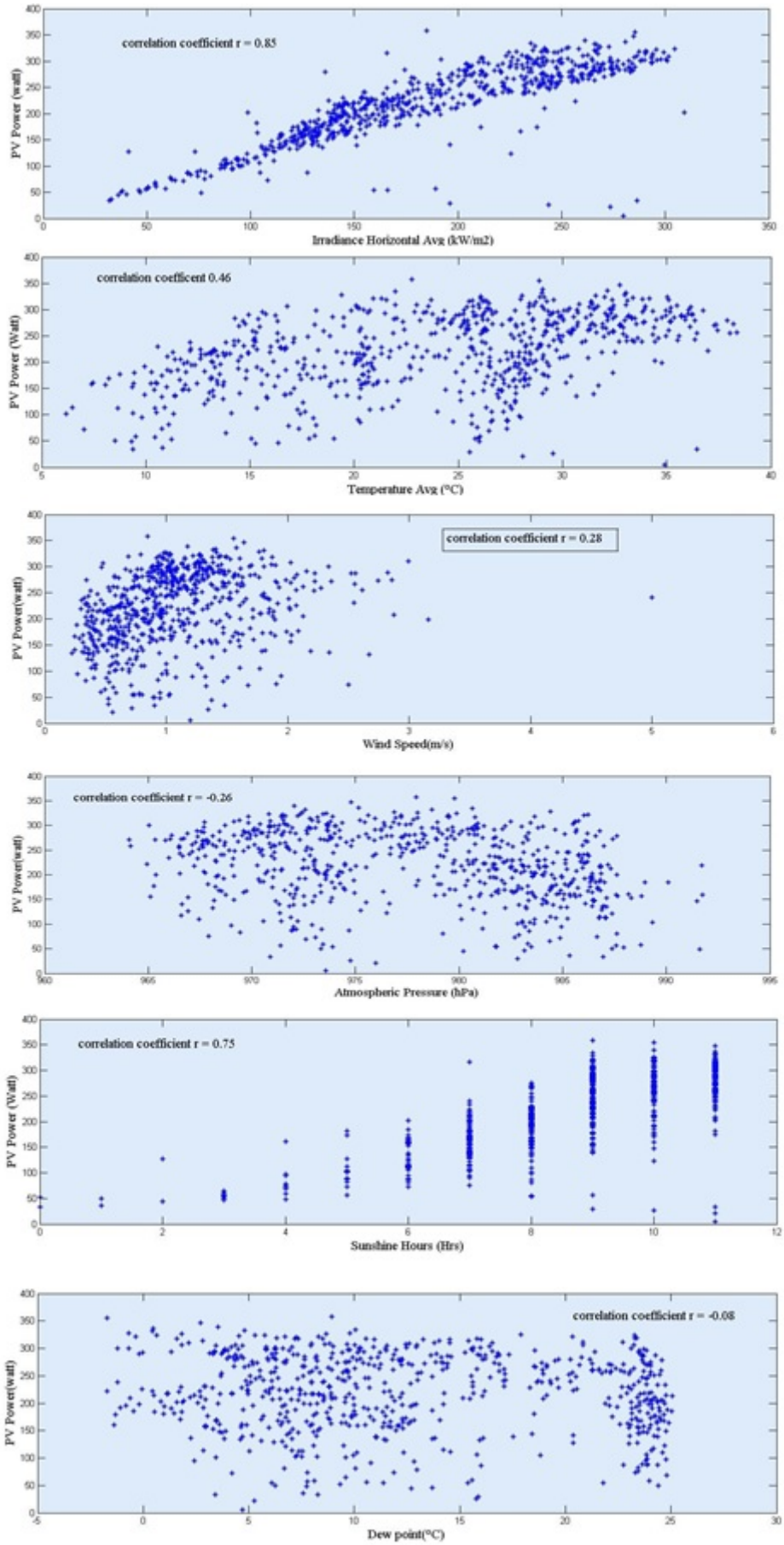


Figure 4.6: Correlation of Meteorological Parameters with Solar Power

4.3 Summary

The data set used for this research consists of meteorological and solar power data. The National Institute of Solar Energy (NISE), situated in Gurguram, was the site's location from where all the data was procured. The data set has a total of 920 observations, which are further split into 620 training samples, 150 validation samples, and 150 testing samples. The meteorological data mainly contains the temperature, pressure, solar radiation, dew point, wind speed, number of days, and sunshine hours. The daily values of each parameter and solar power are calculated using a simple averaging technique.

Chapter 5

Experimental Results

This chapter represents the experimental results obtained during this research work. The solar radiation and solar power estimation results using the SVR model are shown here. The results for degradation computation are also presented in this chapter.

5.1 Introduction

The main objective of this research work is to develop a machine learning model for solar energy estimation. In this context, we first tried to estimate solar radiation. The SVR model is considered for radiation and solar power estimation. The meteorological data collected from the NISE Gurugram is considered an input vector to the SVR model. These parameters are used in different input combinations to form a set. These sets are used to estimate solar radiation and solar power. SVR model's performance is evaluated till we get the best input combination. Once the best combination is known, the SVR model is used to estimate the solar radiation and solar power using test data.

5.2 Solar Radiation Estimation

The most important parameter for solar energy/power estimation is solar radiation, which is not readily available at the site. Therefore, in the first phase of this research problem, we estimated solar radiation via machine learning methods using readily available meteorological parameters. The different input combinations of meteorological parameters are considered to form a set applied to the Support Vector Regression (SVR). The SVR model is trained using 620 training samples out of 920 samples collected from NISE. Table 5.1 represents the sets of input combinations considered as the input to machine learning techniques. Different kernel functions of SVR, like

Table 5.1: Different Combination of Input Parameters

Set Number	Set of Input Parameters				
Set-1	Day	Temperature	Humidity	Pressure	Wind Speed
Set-2	Day	Pressure	Humidity	Wind Speed	
Set-3	Day	Temperature	Pressure	Wind Speed	
Set-4	Day	Temperature	Humidity	Wind Speed	
Set-5	Day	Temperature	Humidity	Pressure	

RBF kernel, linear, quadratic, and cubic regression, are considered. The RBF kernel's performance was better than other kernel functions. Therefore SVR model with RBF kernel function is considered here. Four machine learning techniques, ANN, SVR, HMM, and ANFIS models, were compared on the input sets described in Table 5.1. RMSE error is used to measure the models' performance. RMSE indicates how precise the model is by calculating the difference between the estimated value and the actual value of solar radiation. Table 5.2 and Figure 5.3 demonstrate that the SVR approach produces better results than ANN, HMM, and ANFIS. Temperature, air pressure, relative humidity, time of day, and wind speed have been the most crucial parameters. With all these input parameters in Set-1, the best performance is reached. Table 5.2 presents a comparison of the RMSE to many other methodologies. Figures 5.1 display the estimation results for the actual value for set-1, and Figure 5.2 shows the regression line for testing data.

5.2.1 Discussion on Solar Radiation Estimation Results

The primary conclusions that can be drawn after carefully examining the results are as follows. When Set-1, which includes the number of days, temperature, relative humidity, pressure, and wind speed, is used as input, the SVR model performs better. Compared to other methodologies, this set's RMSE error value, which is 14.3, is much lower. Additionally, it was discovered that temperature is one of the most crucial factors to consider when estimating solar radiation.

Table 5.2: Comparison of RMSE with other Models

Set Number	SVR	ANN	HMM	ANFIS
Set-1	14.3	33.5	20.1	21.3
Set-2	15.7	32.8	22.9	23.3
Set-3	20.1	34.1	26.9	28.3
Set-4	17.1	32.8	22.8	23.1
Set-5	18.7	32.5	23.8	24.7

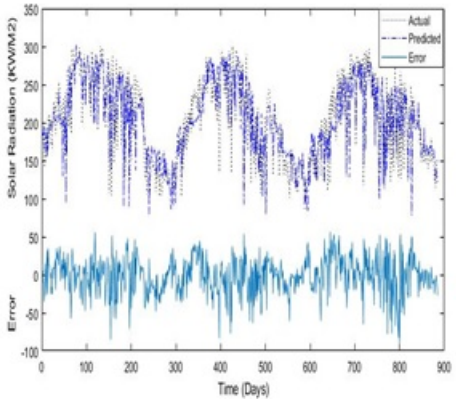


Figure 5.1: Solar Radiation Estimation for Set 1

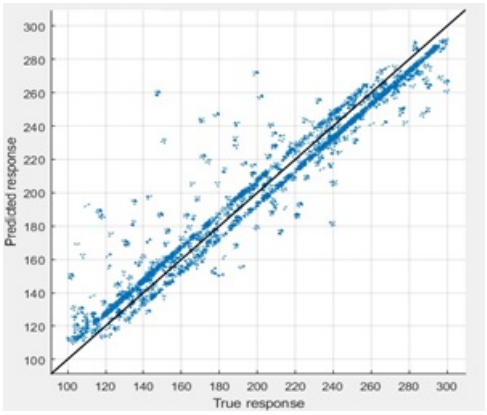


Figure 5.2: Regession line for RBF SVR for Set 1

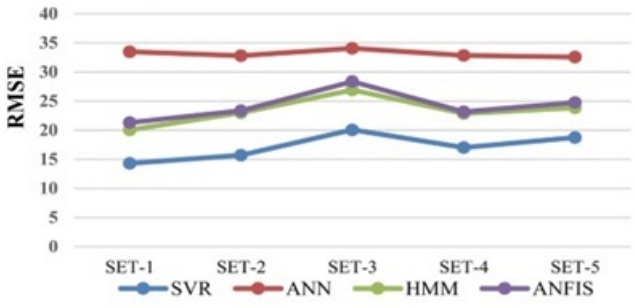


Figure 5.3: RMSE Comparison

5.3 Solar Power Estimation

This section presents the results for solar power prediction. In the previous section, estimation is done for solar radiation, and it was found that the results of the SVR model are better than other machine learning techniques. Therefore, the SVR technique is applied for solar power estimation. However, in this task, we have used solar radiation as one of the input parameters along with sunshine hours which were not used in the case of solar radiation prediction. Hence, seven meteorological parameters are used as an input vector to the SVR model. it is difficult to choose which are essential and which are least important among the seven meteorological parameters. To overcome this difficulty, we have tried 52 input combinations shown in Table 5.3.

Table 5.3: SVR Models with Different Inputs

Models	Inputs	Models	Inputs	Models	Inputs	Models	Inputs
SVR_1	T_{avg}	SVR_{14}	N, D_{pt}	SVR_{27}	$T_{avg}, P_{avg}, R_H, U, D_{pt}$	SVR_{40}	$T_{avg}, P_{avg}, R_H, U, H_0$
SVR_2	H_0	SVR_{15}	H_0, S_t	SVR_{28}	N, T_{avg}, H_0, U, S_t	SVR_{41}	$R_H, U, P_{avg}, H_0, S_t$
SVR_3	S_t	SVR_{16}	T_{avg}, S_t	SVR_{29}	T_{avg}, H_0, R_H, S_t	SVR_{42}	$N, T_{avg}, D_{pt}, U, H_0, S_t$
SVR_4	R_H	SVR_{17}	U, R_H	SVR_{30}	N, P_{avg}, H_0, U, S_t	SVR_{43}	$N, T_{avg}, R_H, U, P_{avg}, H_0$
SVR_5	U	SVR_{18}	T_{avg}, R_H	SVR_{31}	N, H_0, R_H, S_t	SVR_{44}	$N, R_H, U, P_{avg}, H_0, S_t$
SVR_6	P_{avg}	SVR_{19}	T_{avg}, U	SVR_{32}	$T_{avg}, P_{avg}, H_0, S_t$	SVR_{45}	$T_{avg}, R_H, U, P_{avg}, H_0, S_t$
SVR_7	D_{pt}	SVR_{20}	P_{avg}, U	SVR_{33}	T_{avg}, S_t, H_0, S_t	SVR_{46}	$AllParameters$
SVR_8	N	SVR_{21}	N, T_{avg}, S_t	SVR_{34}	N, P_{avg}, H_0, S_t	SVR_{47}	N, H_0
SVR_9	N, T_{avg}	SVR_{22}	N, U, R_H	SVR_{35}	$N, T_{avg}, H_0, R_H, S_t, PR$	SVR_{48}	N, H_0, S_t, PR
SVR_{10}	N, S_t	SVR_{23}	N, T_{avg}, R_H	SVR_{36}	N, T_{avg}, H_0, S_t, PR	SVR_{49}	H_0, T_{avg}, S_t
SVR_{11}	N, R_H	SVR_{24}	$N, _{avg}, U$	SVR_{37}	$N, T_{avg}, H_0, P_{avg}, R_H, S_t, PR$	SVR_{50}	P_{avg}, H_0, S_t
SVR_{12}	N, U	SVR_{25}	N, P_{avg}, R_U	SVR_{38}	$N, T_{avg}, H_0, D_{pt}, S_t$	SVR_{51}	H_0, T_{avg}
SVR_{13}	N, P_{avg}	SVR_{26}	N, T_{avg}, H_0, U	SVR_{39}	$T_{avg}, D_{pt}, U, H_0, S_t$	SVR_{52}	N, H_0, T_{avg}

All 52 models' results are compared based on statistical measures like RMSE, MAPE, RMSE, and R^2 shown in Table 5.4. and Figure 5.4. For a better interpretation of the results, shown in Figure 5.4, we have divided the SVR models into two groups, Group 1 and Group 2, as shown in Figure 5.5. Further, on the basics of the values of the statistical measures, we can categorize the models into three categories, Best models, Average Models, and Worst Models, respectively. The best models are shown in violet color, average models are in blue colors, and red represents the worst models. Table 5.5 is used to represent the statistical measure for best fit models. Once the Best models and their input parameters are identified, the next step is to optimize the SVR parameters C and γ. The SVR parameters C and γ are calculated via the grid search method using 10 cross folds on validation data. 150 data samples are reserved for this process. The C and γ for the best models are calculated, and their values are shown in Table 5.6. The optimized values of C and γ are used further to estimate the

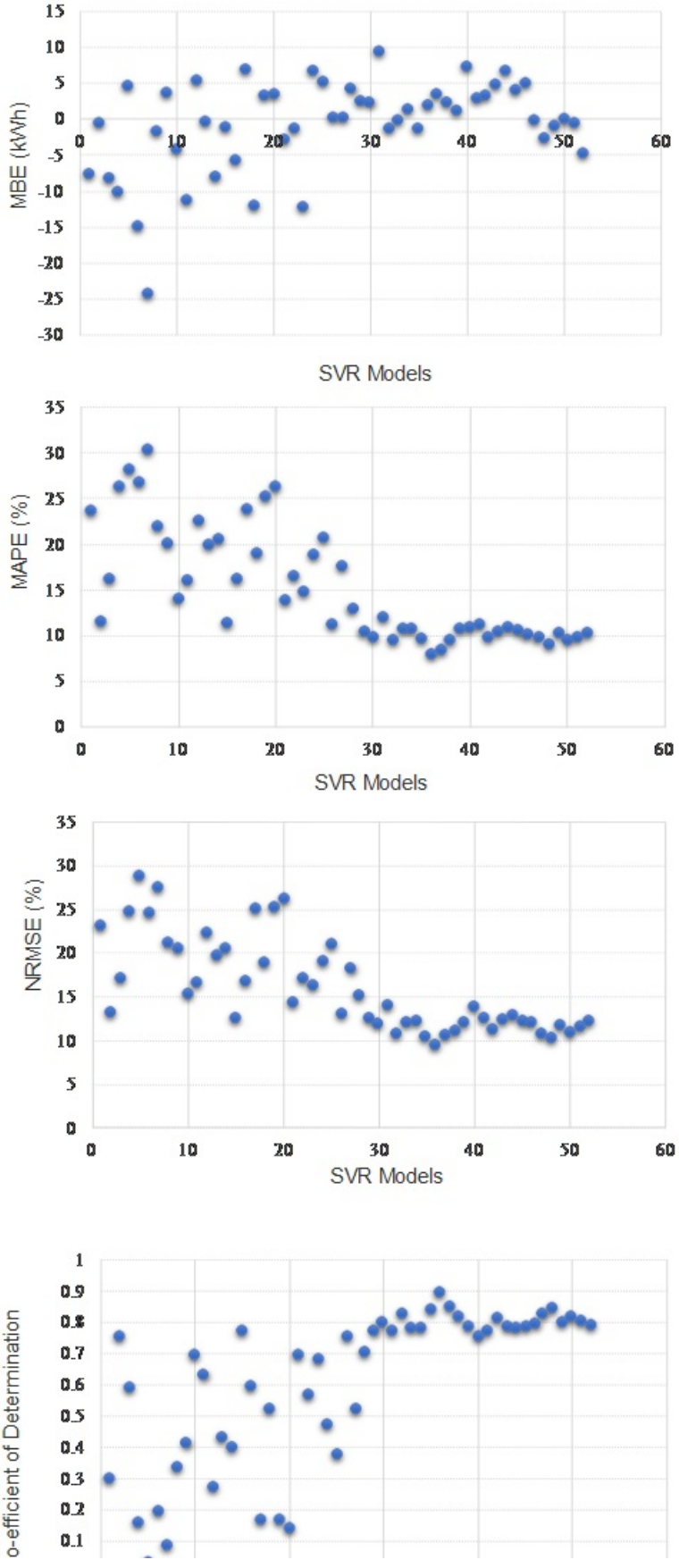


Figure 5.4: MBE, MAPE, RMSE, and R^2 of SVR Models

solar power using testing data. The estimation of the best models for solar power is shown in Figure 5.6.

Table 5.4: Performance Analysis of SVR Models

Models	R^2	NRMSE	RMSE	MAPE	Models	R^2	NRMSE	RMSE	MAPE
SVR_1	0.30	23.17	23.65	-7.56	SVR_{27}	0.52	18.33	17.68	0.19
SVR_2	0.75	13.31	11.57	-0.48	SVR_{28}	0.70	15.26	12.94	4.38
SVR_3	0.59	17.20	16.16	-8.16	SVR_{29}	0.77	12.55	10.41	2.57
SVR_4	0.15	24.86	26.25	-10.13	SVR_{30}	0.80	11.98	9.82	2.29
SVR_5	0.03	28.82	28.12	4.72	SVR_{31}	0.77	14.02	12.04	9.50
SVR_6	0.19	24.64	26.79	-14.93	SVR_{32}	0.82	10.86	9.59	-1.26
SVR_7	0.08	27.56	30.39	-24.28	SVR_{33}	0.78	12.12	10.73	-0.04
SVR_8	0.33	21.20	22.02	-1.74	SVR_{34}	0.78	12.23	10.70	1.49
SVR_9	0.41	20.51	20.04	3.70	SVR_{35}	0.84	10.40	9.72	-1.18
SVR_{10}	0.69	15.28	13.97	-4.23	SVR_{36}	0.89	9.56	8.01	2.07
SVR_{11}	0.63	16.71	16.02	-11.16	SVR_{37}	0.84	10.57	8.45	3.53
SVR_{12}	0.27	22.36	22.59	5.38	SVR_{38}	0.81	11.16	9.49	2.46
SVR_{13}	0.42	19.73	19.97	-0.21	SVR_{39}	0.78	12.04	10.68	1.20
SVR_{14}	0.39	20.57	20.59	-7.95	SVR_{40}	0.75	13.87	10.84	7.43
SVR_{15}	0.77	12.59	11.32	-1.10	SVR_{41}	0.77	12.63	11.16	3.01
SVR_{16}	0.59	16.84	16.26	-5.60	SVR_{42}	0.81	11.30	9.78	3.37
SVR_{17}	0.16	25.13	23.90	7.04	SVR_{43}	0.78	12.45	10.43	4.96
SVR_{18}	0.52	18.87	19.06	-12.06	SVR_{44}	0.78	12.88	10.95	6.82
SVR_{19}	0.16	25.21	25.30	3.30	SVR_{45}	0.78	12.21	10.64	4.03
SVR_{20}	0.13	26.20	26.38	3.46	SVR_{46}	0.79	12.18	10.20	5.10
SVR_{21}	0.69	14.44	13.83	-2.89	SVR_{47}	0.82	10.83	9.83	-0.10
SVR_{22}	0.57	17.22	16.57	-1.19	SVR_{48}	0.84	10.31	9.06	-2.66
SVR_{23}	0.68	16.35	14.87	-12.12	SVR_{49}	0.80	11.70	10.25	-0.91
SVR_{24}	0.47	19.14	18.8361	6.7177	SVR_{50}	0.82	11.02	9.5644	0.0887
SVR_{25}	0.37	21.02	20.7241	5.2843	SVR_{51}	0.80	11.63	9.782	-0.499
SVR_{26}	0.75	13.12	11.3041	0.327	SVR_{52}	0.79	12.27	10.2605	-4.7061

Table 5.5: Best Models with Performance Metrics

Best Models	R^2	NRMSE	MAPE	MBE
$N, T_{avg}, R_H, H_0, S_t, (SVR_{35})$	0.84	10.40	9.72	-1.1
$N, T_{avg}, H_0, S_t, (SVR_{36})$	0.89	9.56	8.01	2.07
$N, T_{avg}, P_{avg}, S_t, (SVR_{37})$	0.85	10.57	8.45	3.53
$N, H_0, S_t, (SVR_{48})$	0.84	10.31	9.06	-2.66

5.3.1 Discussion of Solar Power Estimation

Solar power estimation utilizes the SVR method, which has shown better results than other machine learning techniques used for radiation estimation. Considering solar radiation and sunshine hours as additional inputs in the case of solar power estimation, the input parameters vector dimension increased. However, these inputs

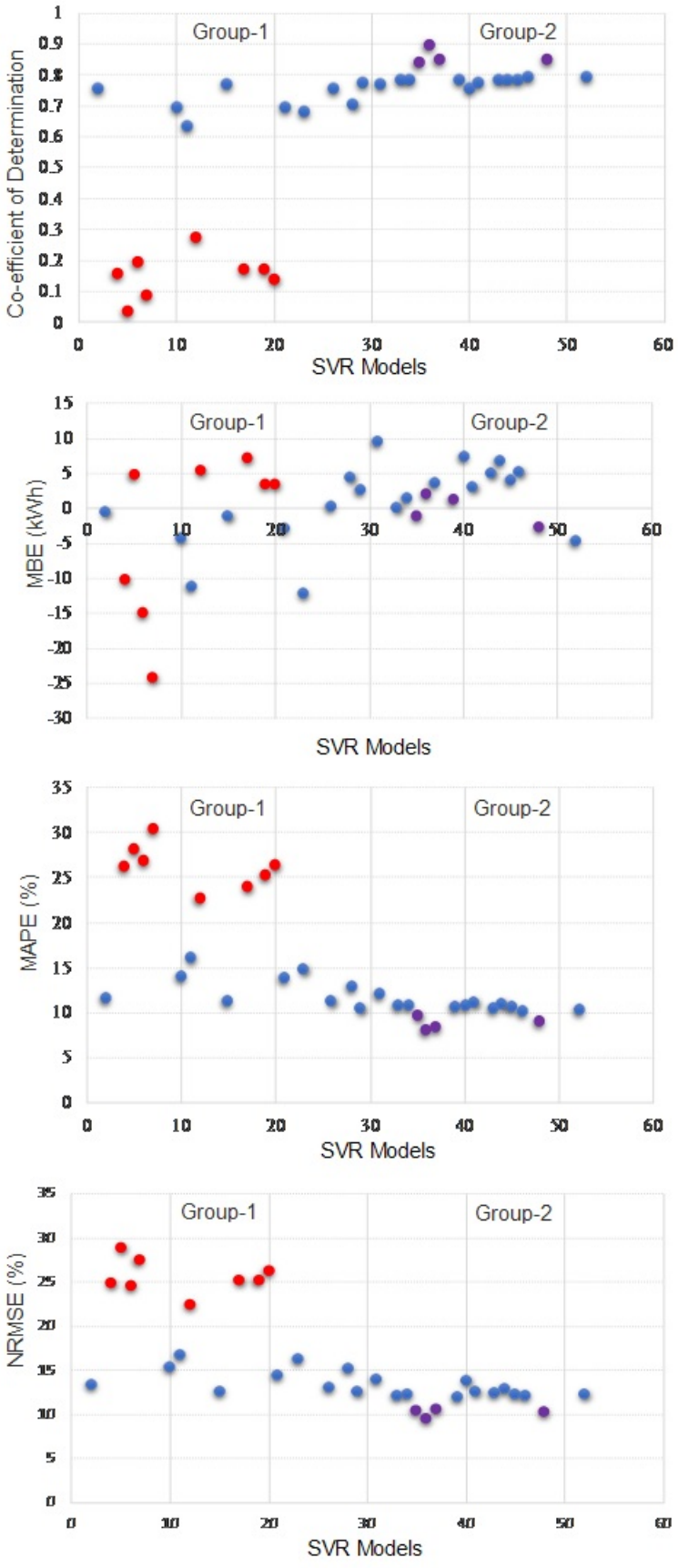


Figure 5.5: Best, Average, and Worst SVR Models

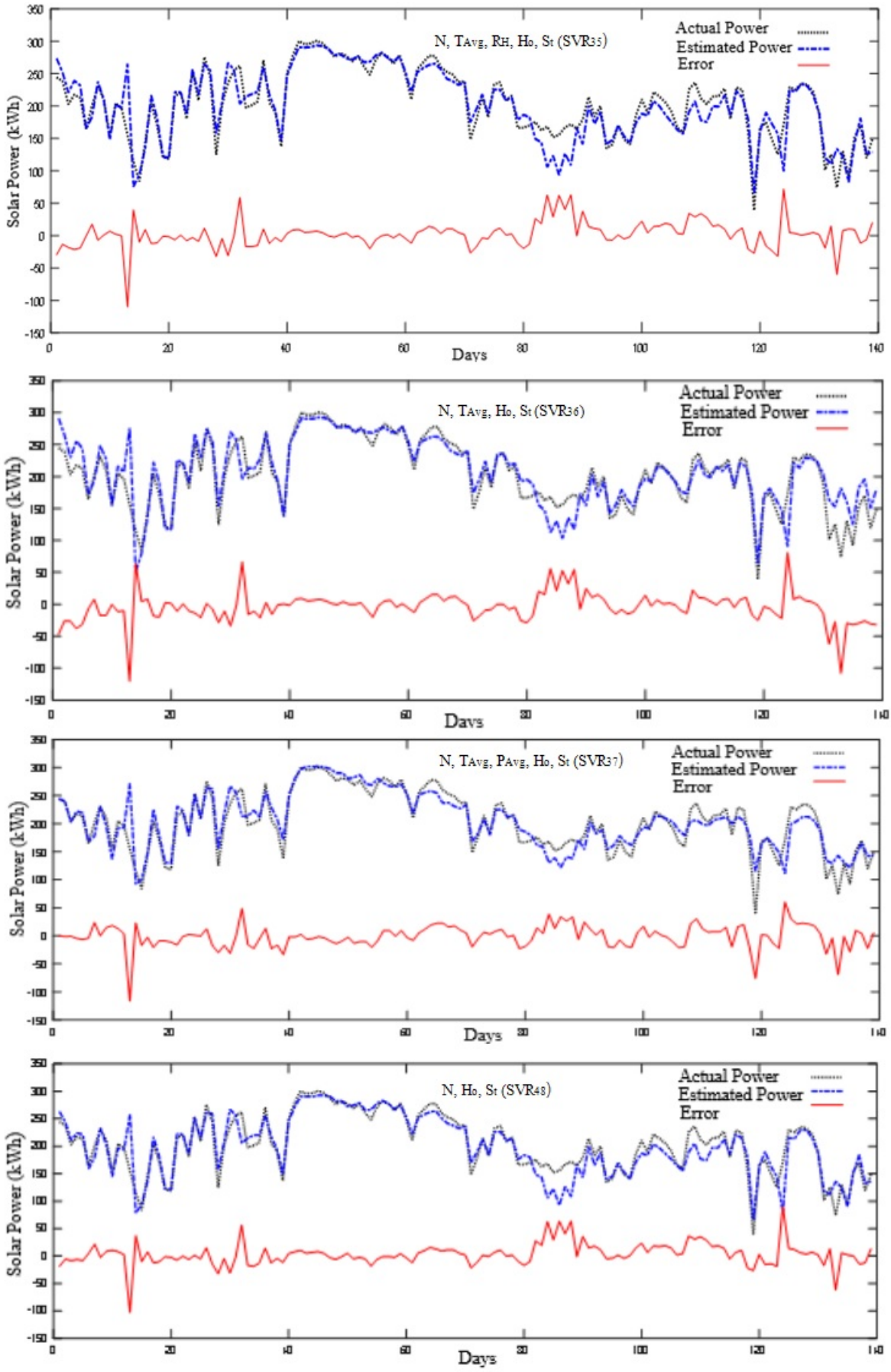


Figure 5.6: Solar Power Estimation of Best SVR Models

Table 5.6: Optimized Values for the Best Models

Best SVR Model	**C**	γ
$N, T_{avg} R_H, H_0, S_t (SVR_{35})$	0.5	8
$N, T_{avg}, H_0, S_t (SVR_{36})$	128	0.125
$N, T_{avg}, P_{avg}, H_0, S_t (SVR_{37})$	128	0.125
$N, H_0, S_t (SVR_{48})$	0.5	2

are the most essential. As a first step, the SVR is used to estimate the solar power from each model ($SVR_1 \ldots SVR_{52}$) having different input combinations of all the seven inputs. It is clear from the performance of different models shown in Figure 5.5 that these models can be dichotomies into Group-1 and Group-2. The Group-2 models are better than the Group-1 models in terms of accuracy. The range of R^2 values for Group-1 and Group-2 are (0.331-0.775) and (0.755-0.896) respectively. The error metrics MBE, NRMSE, and MAPE, are all under 10.54 for Group-1, while this range is higher(greater than 10.54) for the Group-2 models. Out of the 52 models, four models ($SVR_{35}, SVR_{36}, SVR_{37}, SVR_{48}$) have been selected with the highest accuracy in terms of R^2, and the lowest values of error metrics in term of MBE, NRMSE, MAPE as shown in Table-5.4. The most important parameters were solar radiation, sunshine hour, temperature, and the number of days. The SVR_{36} model has all these parameters as its input and has shown superior results among the four best models. The R^2 value reported for this model was 0.89, while its error was also reported as less than 10.

5.4 Computing Degradation of PV Panel

This section is dedicated to the performance monitoring of PV panels. Photovoltaic system performance must be monitored for effective solar energy consumption. One of the important objectives of this research is to find the degradation induced in PV panels. For evaluating the degraded performance of PV panels, this research suggests a model based on the clustering technique discussed in chapter 3. The proposed model is helpful for real-time estimation of the PV panels' performance ratio (PR) and, as a result, more accurate forecasting of PV power production. The segmental K-means clustering technique determines clusters of input meteorological data with similar input conditions. The suggested model takes the meteorological data as input. In this

research work, we will evaluate the performance ratio and degradation rate in output solar power in terms of PR for panels with three different topologies over three years: hetero-junction with an intrinsic thin layer (HIT), poly-crystalline silicon (p-Si), and amorphous silicon (a-Si). The results obtained for the degradation rate for three PV technologies with the help of CCDR are shown in Figure 5.7, and their comparison with other methods is presented in Table 5.5. The results are compared with similar studies conducted in USA and Cyprus. The process of PR calculation starts by assigning the cluster number to each meteorological data. For a better understanding of the process, let us consider a-Si PV panels. Once the cluster number is known, the next step is to find the common monthly clusters in each year from 2010-2012. Once the common monthly cluster is identified next step is to calculate the monthly average solar power of common clusters. The monthly average solar power helps to find the PR ratio easily. The monthly average solar power and PR for a-Si are shown in Figure 5.7. The same process is applied to the remaining PV panel technologies. The degradation rate is calculated from the PR, derived via the CCDR method. PR of hetero-junction with an intrinsic thin layer, poly-crystalline silicon, and amorphous silicon is shown in Figure 5.8. The comparison between the three technologies based on the PR is shown in Figure 5.9. Finally, the degradation rates for the three separate technology modules are computed using the PRs obtained by the CCDR method. The conventional least-square regression method is used to calculate the degradation rates. In figure 5.10, the degradation rate is determined by the slope of the regression line from the monthly PR.

Table 5.7: Comparison of CCDR Result with other Methods

Time frame	PV Panels	Technique	Degradation Rate	Location
Monthly	Mono-silicon	PR based	0.84	USA
	Poly-silicon		0.95	
	HIT		0.95	
Monthly	Crystalline-silicon	PR-LLS	0.63	Cyprus
	HIT		2.03	
Monthly	Amorphous-silicon	PR based	0.85	NISE Gurgugram
	Poly-silicon		0.95	
	HIT		1.1	

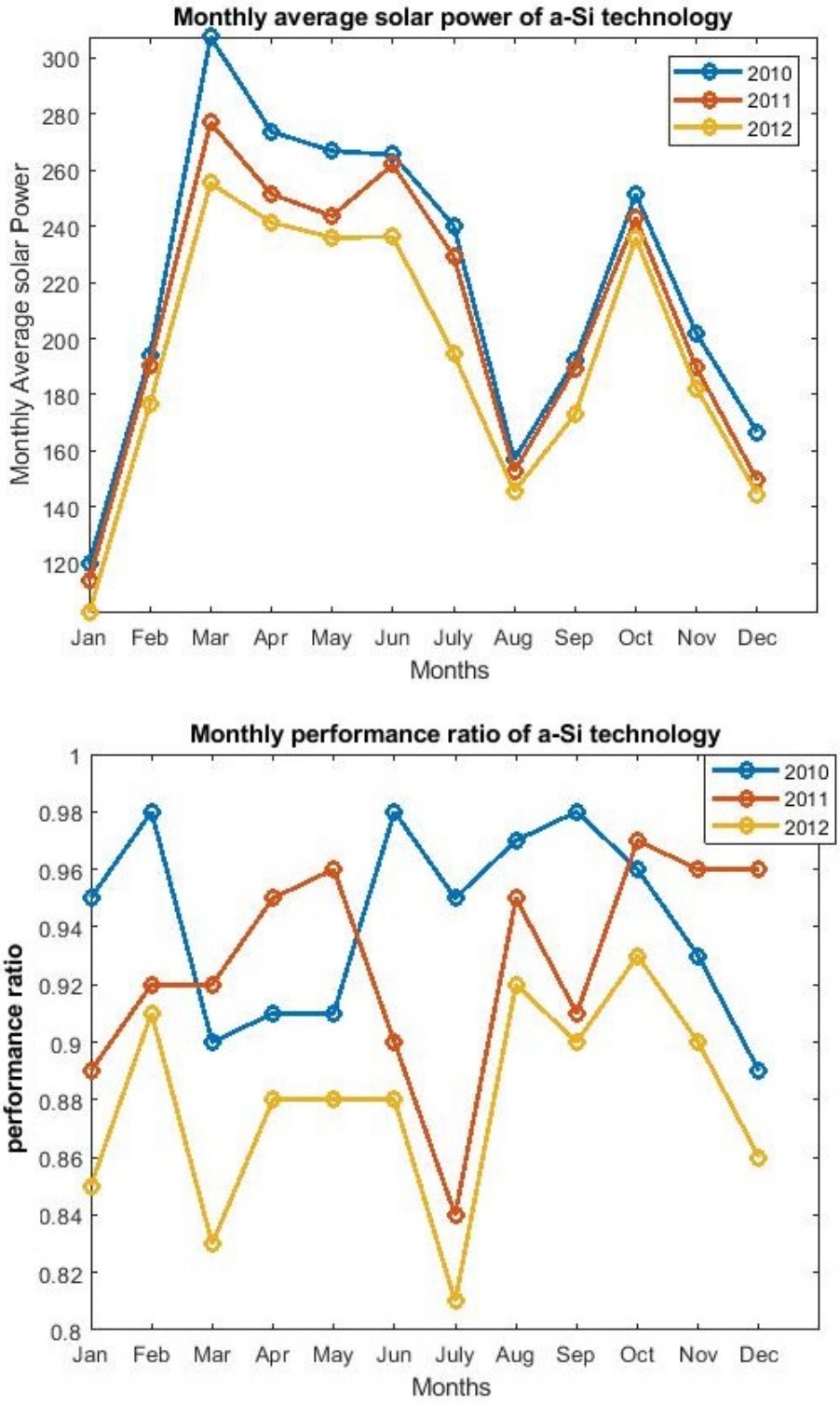


Figure 5.7: Average Solar Power & PR of a-Si PV Panels

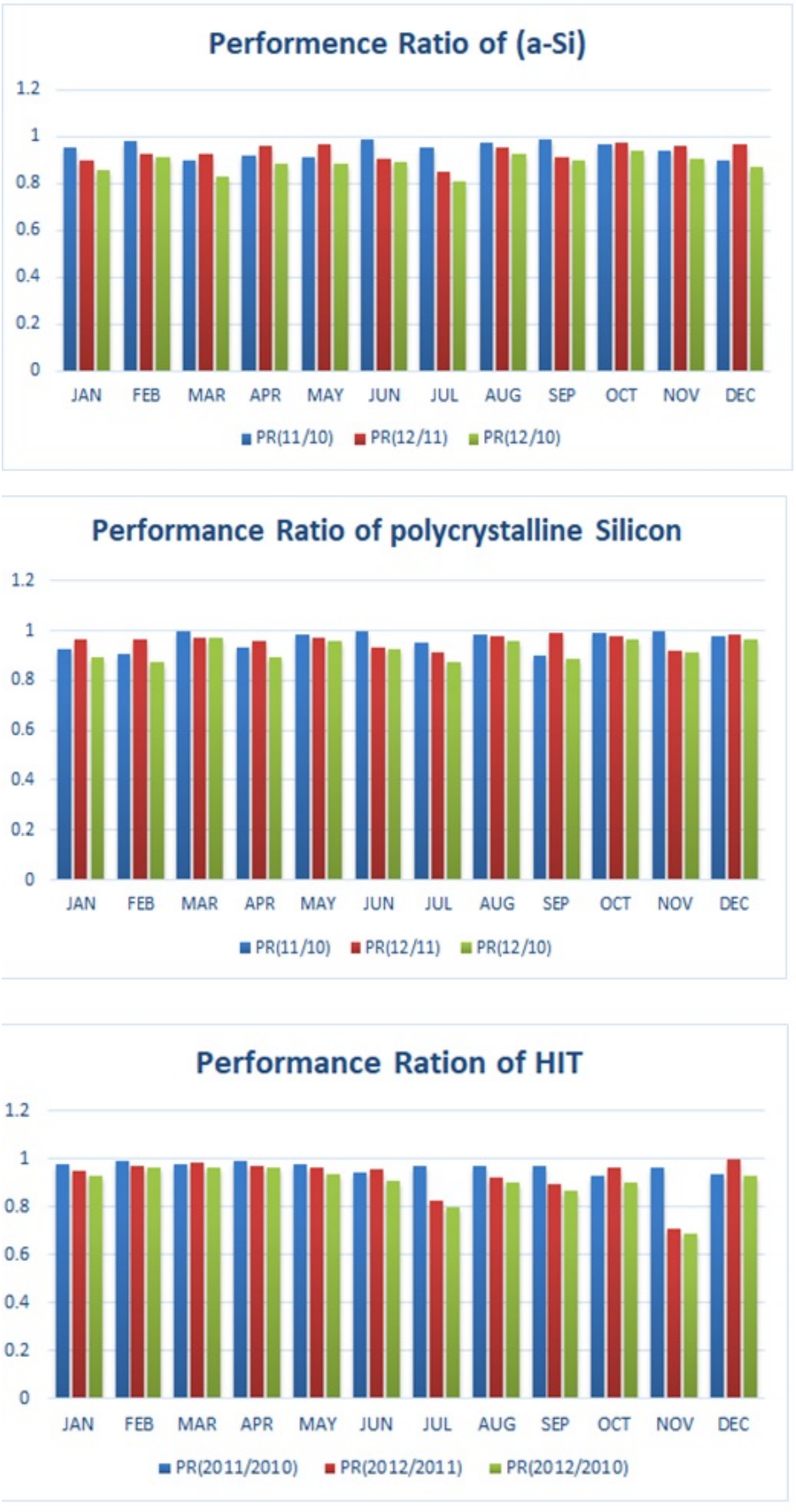


Figure 5.8: PR of a-Si, p-Si, and HIT PV Panels

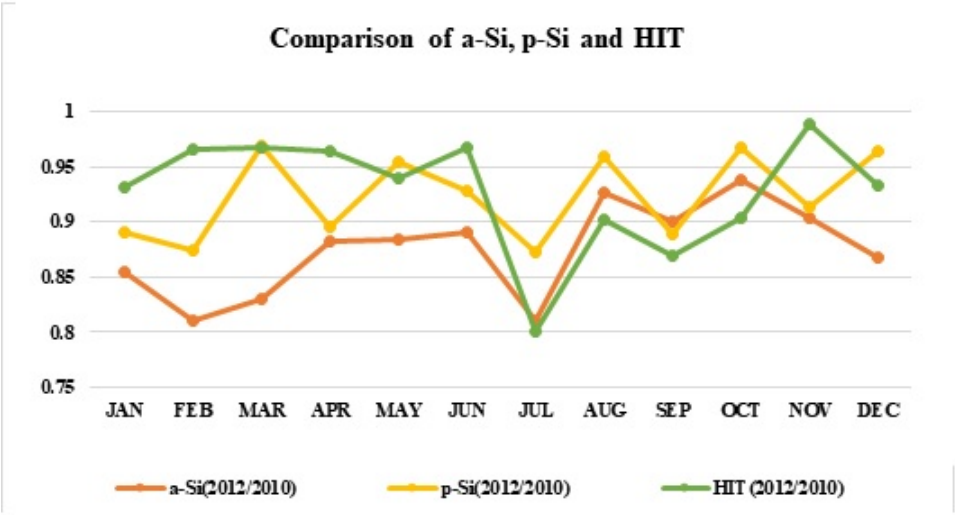


Figure 5.9: Performance Comparison of a-Si, p-Si and HIT Technology

5.4.1 Discussion on CCDR Results

The following conclusion is drawn from the results:

The a-Si technology showed consistently less PR values in the year 2012 when compared with the power produced in 2010, as indicated by PR values consistently less than 1. The PR (2012/2010) values varied between 0.80 and 0.93, with the minimum value in July and the maximum value in October, as shown in Figure 5.8. The average value of PR (2011/2010) was 0.89, which dropped to 0.87 for (2012/2010). Further, the mean degradation rate reported was 0.85% per year for the a-Si technology, calculated by the regression method, as shown in Table 5.7.

The p-Si technology achieved a better PR as compared with the a-Si technology. The PR (2012/2010) values varied between 0.87 and 0.96, with a minimum in July and a maximum value in December, as shown in Figure 5.8. The average value reported for PR (2011/2010) was 0.95, which dropped to 0.91 for (2012/2010), representing a lower extent of degradation compared with the a-Si technology. The mean degradation rate for p-Si technology was found to be 0.95% per year, slightly higher than the a-Si technology as per Table 5.7.

The HIT technology achieved PR (2012/2010) values between 0.80 and 0.97, with the minimum value in July and the maximum value reported in November, as shown in

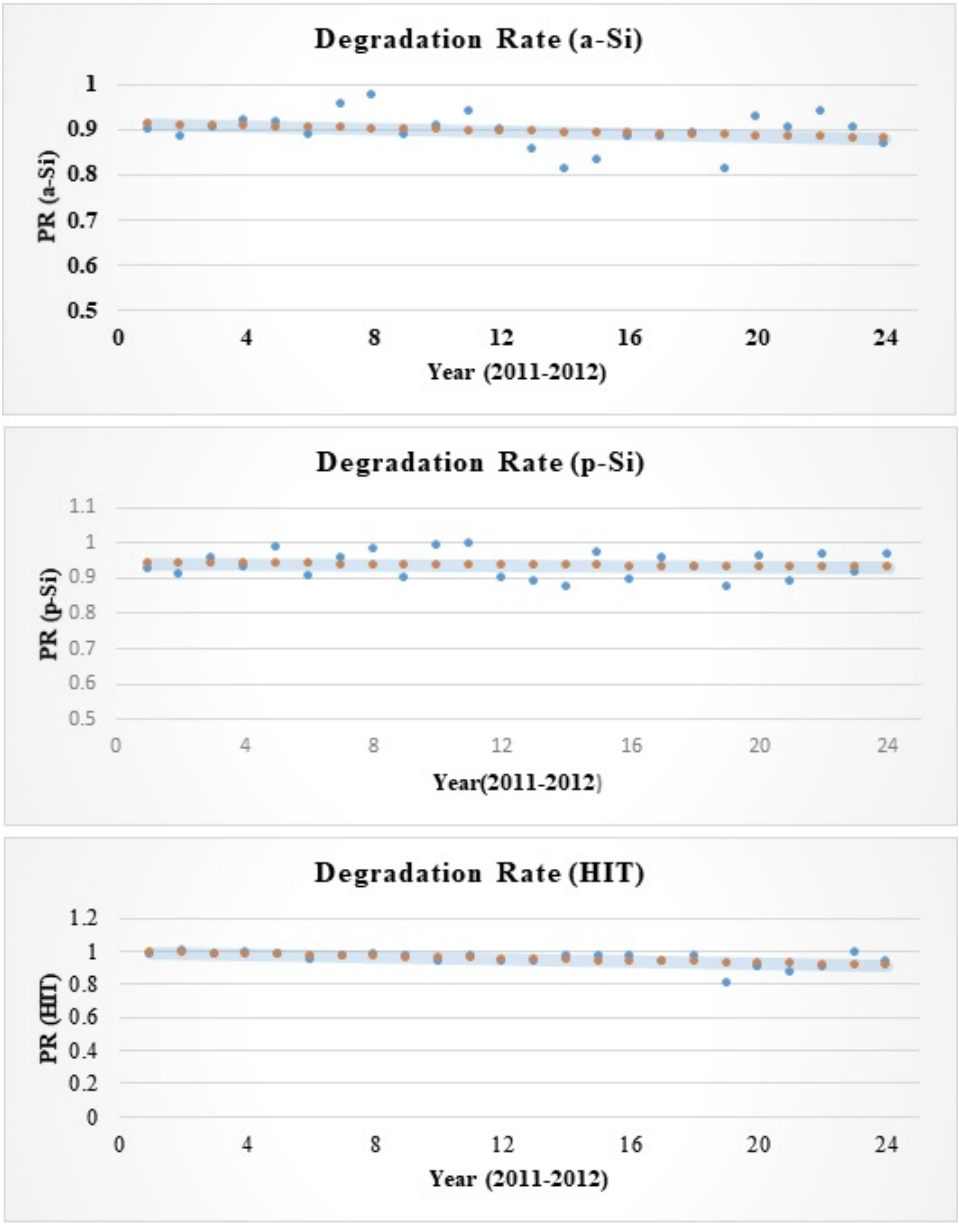


Figure 5.10: Degradation Rate of a-Si, p-Si and HIT Technology PV Modules

Figure 5.8. The average value for PR (2011/2010) was found to be 0.96, which decreased to 0.92 for (2012/2010). The HIT technology showed the highest degradation rate, with a value of 1.1% per year, as shown in Table 5.7.
A comparison of the three technologies is shown in Figure 5.9. The HIT technology performed better than a-Si and p-Si for the year's first half, except for May. The p-Si technology performed better for the year's second half, except for November.
The HIT technology showed the highest degradation rate, with a value of 1.1% per year, followed by p-Si and a-Si technology. A comparison with other degradation rate estimations is shown in Table 5.7.

5.5 Summary

This chapter presented the results obtained for solar radiation and solar power estimation using the SVR model. The SVR model is trained for prediction work using the different combinations of meteorological parameters as the input vector. The performance of each model was evaluated using statistical measures like MAE, MAPE, NRMSE, and R^2. The best-performing model was further used to estimate the test data input. The temperature, Number of days are the crucial parameters for solar radiation estimation. The inputs that contribute effectively in the case of solar power estimation were solar radiation, sunshine hour, temperature, and the number of days. A total of 52 models were trained for solar power estimation work, and four models' performance was found up to the mark. They are further used to estimate solar power for test data input. The solar power output is further investigated for the degradation induced in PV panels. The CCDR method helps us to find the degradation in PV panels. The a-Si panels have less degradation rate, while the HIT panels show the highest degradation rate as per the results obtained with CCDR in Table 5.7.

Our next task is to use this degradation information in estimation to fine-tune the estimation results further and to improve estimation accuracy. This part of the research is presented in chapter 6 of this thesis.

Chapter 6

Real-Time Estimation of Solar Power

This chapter explores the real-time estimation of solar power using the degradation information obtained in chapter 5 for estimating the solar power of PV panels. Estimation accuracy increased significantly by considering the degradation induced in PV panels.

6.1 Introduction

The real-time solar power estimation explores the possibility of including the degradation information in the estimation work. In this research, we have estimated solar power and found the degradation induced in the PV panels using machine learning techniques (SVR and CCDR) presented in chapters 3 and 4. The estimation accuracy can be further increased if we include the degradation information in the estimation work. The complete real-time estimation process includes combining all the steps starting from the data collection, data pre-processing, estimating the solar power, computing the degradation, and then adding the degradation results in estimation work.

6.2 Methodology for Real-time Estimation

The methodology for the real-time estimation and refinement of the solar power in sync with the degradation in PV panels is broadly divided into four modules, as shown in Figure 6.1.

6.2.1 Module-1

This section contains the pre-processing of meteorological data, detailed in chapter 3 (Data and Site Description). The meteorological data have different ranges, so it is normalized, cleaned, and converted in the required format in module 1. Both machine learning techniques, SVR and CCDR, utilize the data received from module 1.

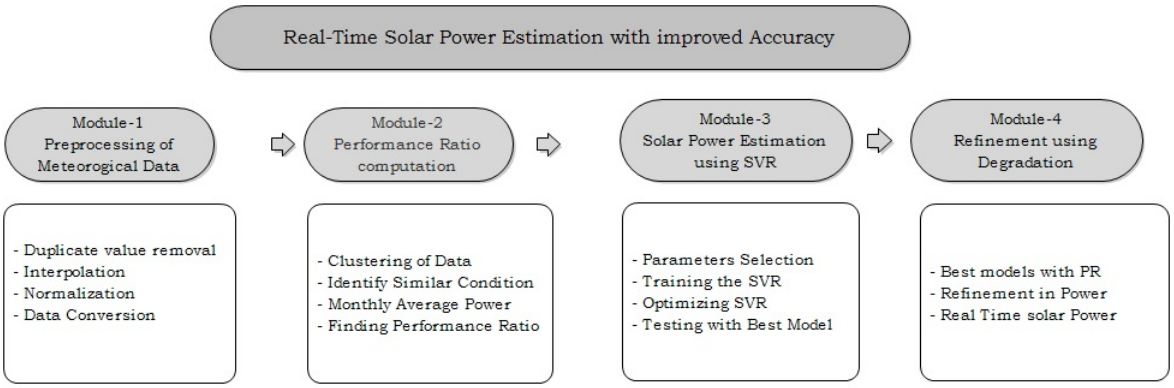


Figure 6.1: Methodology for Real-Time Estimation of Solar Power

6.2.2 Module-2

This module performs the task of computing the PV panel's degradation information in terms of performance ratio (PR). The CCDR method computes the degradation based on the clustering technique. Similar weather conditions in meteorological parameters were obtained via k-means clustering, and after that, the PR and degradation were calculated as per the process discussed in chapter 3 (CCDR).

6.2.3 Module-3

Module 3 is dedicated to the prediction work. The SVR models are built for the different inputs of meteorological data. The best models are selected from them to use further in real-time estimation. Four SVR models are selected for this task, as discussed in chapter 5(Solar Power Estimation Results). Now, we will use these models to find real-time solar power.

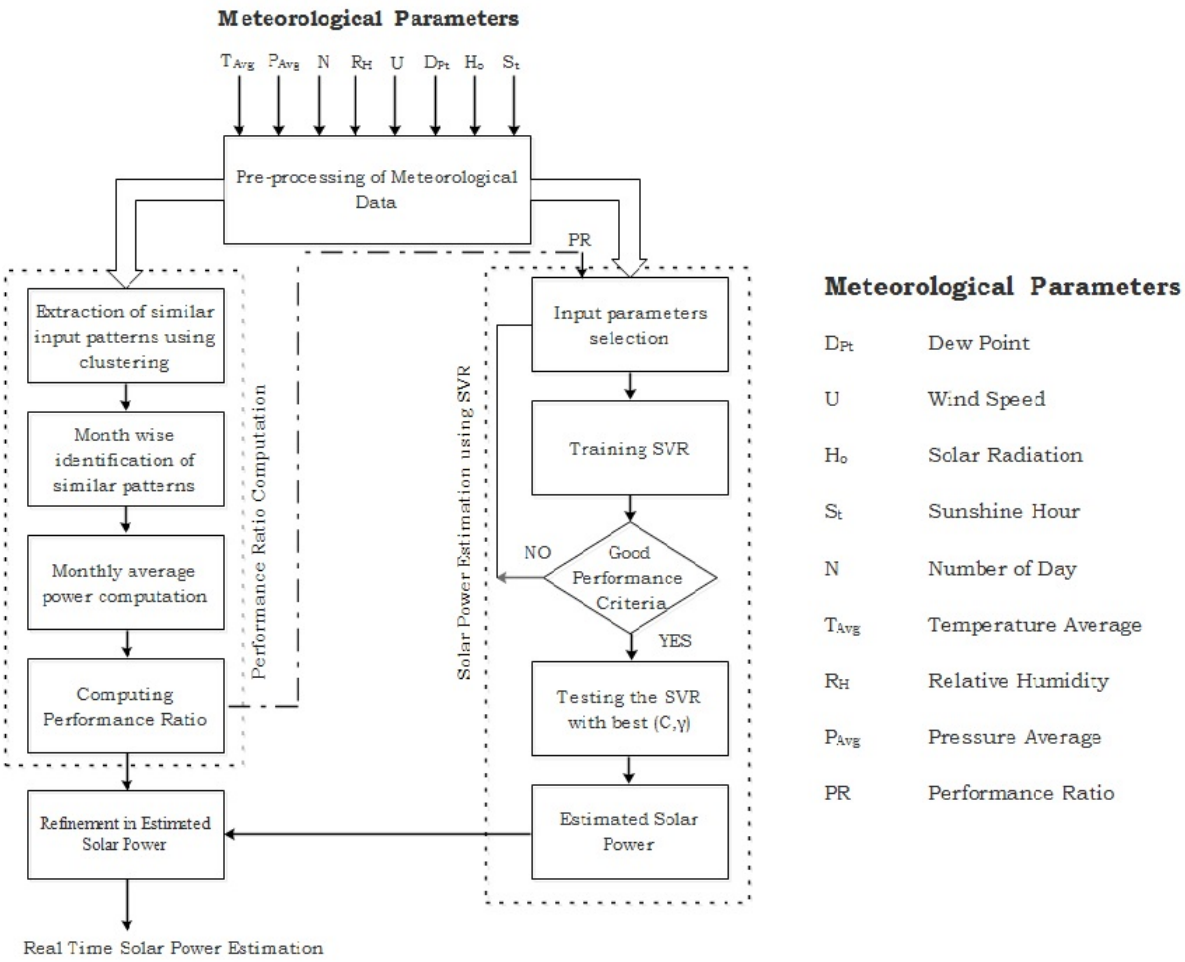


Figure 6.2: Flowchart for Real-Time Estimation of Solar Power

6.2.4 Module-4

The real-time solar power estimation is achieved in module 4 by considering the information of degradation induced in PV panels during the estimation. One of the methods to achieve it is to consider the PR as one of the input parameters for the SVR model, as shown in Figure 6.2. The PR is calculated via the CCDR method for each PV technology. Finally, PR and the meteorological parameters are applied to the SVR model. We have considered the effect of PR on best-performing models. The results are compared for the best models using statistical measures with and without PR as input. The addition of the performance ratio factor with meteorological parameters of the four best-performing models has shown a significant improvement in terms of accuracy for the power estimation. However, another experiment was performed to reduce computational complexity and time. The solar power output was refined by attenuating the model output as per the performance ratio factor during that period,

as shown in Equation 6.2.1.

$$Refined\ Power = Estimated\ Power * PR \tag{6.2.1}$$

After analyzing the results of both approaches, it is concluded that the experimental results from the first approach are slightly better than the second one; however, the latter has an additional advantage of reduced computational complexity and time. The results of the best models are compared before and after applying the PR correction. The results have shown a significant improvement.

Table 6.1: Best Models with Performance Metrics

Best models performance	R^2	**NRMSE**	**MAPE**	**MBE**
$N, T_{avg}, R_H, H_0, S_t, (SVR_{35})$	0.84	10.40	9.72	-1.1
$N, T_{avg}, H_0, S_t, (SVR_{36})$	0.89	9.56	8.01	2.07
$N, T_{avg}, P_{avg}, S_t, (SVR_{37})$	0.85	10.57	8.45	3.53
$N, H_0, S_t, (SVR_{48})$	0.84	10.31	9.06	-2.66

Table 6.2: Performance Statistics of Best Models After Refinement Process

Best models performance	R^2		**NRMSE**		**MAPE**		**RMSE**	
	Before	***After***	***Before***	***After***	***Before***	***After***	***Before***	***After***
$N, T_{avg}, R_H, H_0, S_t, (SVR_{35})$	0.84	0.90	10.40	8.12	9.72	7.46	21.19	16.56
$N, T_{avg}, H_0, S_t, (SVR_{36})$	0.89	0.92	9.56	7.10	8.01	5.53	21.54	14.50
$N, T_{avg}, P_{avg}, S_t, (SVR_{37})$	0.85	0.93	10.57	6.45	8.45	5.17	21.64	13.15
$N, H_0, S_t, (SVR_{48})$	0.84	0.93	10.31	6.85	9.06	5.53	21.01	13.97

6.3 Discussion on Real-time Estimation Results

The best fit models ($SVR_{35}, SVR_{36}, SVR_{37}, SVR_{48}$) are applied with PR input has shown the significant improvement in their results. The coefficient of determination R^2 increased from 0.84 to 0.94. The error metrics NRMSE, MAPE, and RMSE were also reduced from 10.57 to 6.45, 9.72 to 5.17, and 21.54 to 13.15. The estimation curve also shows a significant improvement, as shown in Figure 6.3.

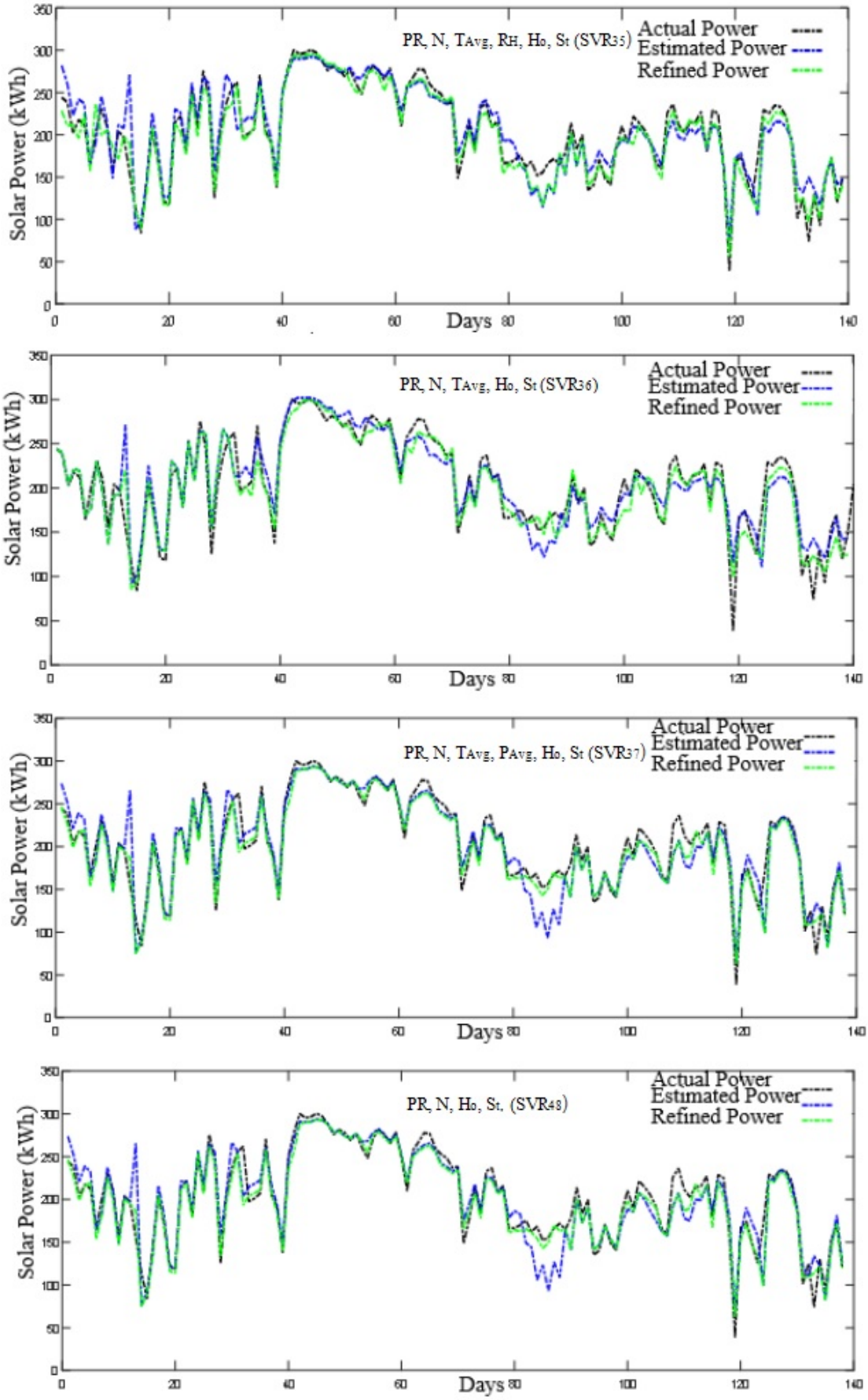


Figure 6.3: Real-time Estimated Solar Power for Best Models

Chapter 7

Conclusions and Future Directions

The research work presented in this thesis is mainly used for estimating solar power from PV panels in real time. The degradation parameter is included in the estimation to improve the accuracy of estimation.

7.1 Conclusions

The research work embodied in this thesis has presented a new machine learning-based technique for computing the degradation in PV panels, "Clustering-based Computation of Degradation Rate" (CCDR), which is further utilized to improve solar energy estimation accuracy. The method developed can compute the degradation of PV panels in real time without an on-site inspection. Therefore, it is possible to monitor the performance of the PV plant from a remote location. The model requires the historical meteorological data and solar power data of the plant. As the developed method eliminates the requirement of on-site imaging and physical inspection hence, it reduces the cost of manpower and equipment.

The degradation of three technologies PV panels: amorphous silicon (a-Si), poly-crystalline silicon (p-Si), and the heterojunction with thin intrinsic layer (HIT) was calculated in this thesis. The a-Si PV panels have shown less degradation, followed by poly-crystalline silicon (p-Si) and the hetero-junction with a thin intrinsic layer (HIT) PV panels. The HIT panels degrade at a faster rate.

After calculating the degradation using the CCDR technique, the next step is to incorporate the effect of degradation in terms of performance ratio (PR) to enhance the estimation accuracy of real-time solar power. The PR values obtained via CCDR are applied as one of the inputs to the SVR model. The results showed significant improvement after incorporating the degradation information in the estimation. The developed models are further validated on real-time data set of three different technol-

ogy panels. Thus this research work contributes to a new method based on machine learning techniques for solar energy estimation for PV panels.

7.2 Scope for future study

Many issues in solar power estimation should be addressed and can be worked on in a future extension of this work.

- The present research work can be extended to add the information obtained from sky images taken either from satellite or ground cameras installed at the monitoring station. The information obtained from the images will increase the accuracy of estimation. However, this type of forecast is generally considered in very short-term predictions, also known as now cast. The time horizon for such estimation ranges from 15 minutes to 2 hours. The present work's time horizon is intra days ranging from days to months, which is the case of long-term estimation.

- The proposed technique used the direct estimation method as solar power and radiation are available on site. An indirect approach like the Numerical Weather Prediction (NWP) method can also be used if the radiation data is unavailable. However, it increases the complexity of the model. The NWP method is helpful for locations where meteorological data is not easily available.

- Due to deteriorating weather or poor handling of PV panels, the performance of PV panels degrades. However, it generally occurs in old-aged panels. Sometimes newly installed PV panels also underperform due to poor handling and maintenance. This can be corrected if we keep monitoring the PV panels' performance. Thus the present approach is also helpful for monitoring the performance of PV panels.

- The proposed approaches can be extended for the performance monitoring of concentrated solar plants.

List of Publications

Referred journals:

[1] **P. Bhola** and S. Bhardwaj, "Clustering-based Computation of Degradation rate for photovoltaic technology," ***Journal of Renewable Sustainable Energy***, vol-11, issue-1, pages 014701-9, Jan-2019. **(SCI-indexed)**

[2] **P. Bhola** and S. Bhardwaj, "Accuracy improvement of Solar Power Estimation using Real-Time Degradation computation of PV panels," ***Journal of Circuits, Systems and Computers***, Vol-30, pp- 2150242-246, Oct-2021. **(SCI-indexed)**

[3] **P. Bhola** and S. Bhardwaj, "Estimation of Solar Radiation using Support Vector Regression," ***Journal of Information and Optimization Sciences***, vol-40, issue-2, pp-339-350, March-2019. **(ESCI-indexed)**

International conferences:

[1] **P. Bhola** and S. Bhardwaj, "Solar Energy Estimation Techniques: A Review," ***7th Indian International IEEE Conference on Power Electronics (IICPE)***, Patiala, vol-7, pp-5-9, 2016.

Bibliography

[1] L.-M. Liu, G. B. Hudak, G. E. P. Box, M. E. Muller, and G. C. Tiao, *Forecasting and time series analysis using the SCA statistical system*, vol. 1. 1 ed., 1992.

[2] C. F. C. Hugo T.C. Pedro, "Assessment of forecasting techniques for solar power production with no exogenous inputs Hugo," *Solar Energy*, vol. 86, no. 3-4, pp. 2017–2028, 2012.

[3] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-term spatio-temporal forecasting of photovoltaic power production," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 538–546, 2018.

[4] A. Ahmad, T. N. Anderson, and T. T. Lie, "Hourly global solar irradiation forecasting for New Zealand," *Solar Energy*, vol. 122, pp. 1398–1408, 2015.

[5] V. Prema and K. Uma Rao, "Development of statistical time series models for solar power prediction," *Renewable Energy*, vol. 83, pp. 100–109, 2015.

[6] A. Moreno-Munoz, J. De La Rosa, R. Posadillo, and V. Pallares, "Short term forecasting of solar radiation," in *2008 IEEE International Symposium on Industrial Electronics*, pp. 1537–1541, IEEE, 2008.

[7] V. Kostylev and a. Pavlovski, "Solar Power Forecasting Performance - Towards Industry Standards," *1st Int. Workshop on the Integration of Solar Power into Power Systems*, 2011.

[8] H. T. Pedro, C. F. Coimbra, M. David, and P. Lauret, "Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts," *Renewable Energy*, vol. 123, pp. 191–203, 2018.

[9] I. Ashraf and A. Chandra, "Artificial neural network based models for forecasting electricity generation of grid connected solar pv power plant," *International Journal of Global Energy Issues*, vol. 21, no. 1-2, pp. 119–130, 2004.

[10] B. Singh, D. T. Shahani, and A. K. Verma, "Neural network controlled grid interfaced solar photovoltaic power generation," *IET Power Electronics*, vol. 7, no. 3, pp. 614–626, 2014.

[11] C. Ben Salah and M. Ouali, "Comparison of fuzzy logic and neural network in maximum power point tracker for pv systems," *Electric Power Systems Research*, vol. 81, no. 1, pp. 43–50, 2011.

[12] S. M. Ai and H. A. Ai, "An ANN-Based Approach for Predicting Global Radiation in Locations with No Direct Measurement Instrumentation," *Renewable Energy*, vol. 14, no. 4, pp. 199–204, 1998.

[13] V. Raghuwanshi and S. P. Tiwari, *Flexible Organic Field-Effect Transistors for Biomimetic Applications*, pp. 315–333. Singapore: Springer Singapore, 2022.

[14] G. Rajput, G. Raut, M. Chandra, and S. K. Vishvakarma, "Vlsi implementation of transcendental function hyperbolic tangent for deep neural network accelerators," *Microprocessors and Microsystems*, vol. 84, p. 104270, 2021.

[15] S. J. Fong, G. Li, N. Dey, R. Gonzalez-Crespo, and E. Herrera-Viedma, "Finding an accurate early forecasting model from small dataset: A case of 2019-nCoV novel coronavirus outbreak," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 1, p. 132, 2020.

[16] J. C. Lam, K. K. Wan, and L. Yang, "Solar radiation modelling using ANNs for different climates in China," *Energy Conversion and Management*, vol. 49, no. 5, pp. 1080–1090, 2008.

[17] F. Almonacid, P. Pérez-Higueras, E. F. Fernández, and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a pv generator," *Energy Conversion and Management*, vol. 85, pp. 389–398, 2014.

[18] K. Dahmani, R. Dizene, G. Notton, C. Paoli, C. Voyant, and M. Laure, "Estimation of 5-min time-step data of tilted solar global irradiation using ANN (Artificial Neural Network) model," *Energy*, vol. 70, pp. 374–381, 2014.

[19] T. T. Teo, T. Logenthiran, and W. L. Woo, "Forecasting of photovoltaic power using extreme learning machine," in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, pp. 1–6, 2015.

[20] G. Notton, C. Paoli, L. Ivanova, S. Vasileva, and M. L. Nivet, "Neural network approach to estimate 10-min solar global irradiation values on tilted planes," *Renewable Energy*, vol. 50, pp. 576–584, 2013.

[21] A. Yona, T. Senjyu, T. Funabashi, and C. H. Kim, "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 527–533, 2013.

[22] S. Rehman and M. Mohandes, "Artificial neural network estimation of global solar radiation using air temperature and relative humidity," *Energy Policy*, vol. 36, no. 2, pp. 571–576, 2008.

[23] M. Ozgoren, M. Bilgili, and B. Sahin, "Estimation of global solar radiation using ann over turkey," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5043–5051, 2012.

[24] A. Koca, H. F. Oztop, Y. Varol, and G. O. Koca, "Estimation of solar radiation using artificial neural networks with different input parameters for mediterranean region of anatolia in turkey," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8756–8762, 2011.

[25] R. Azimi, M. Ghayekhloo, and M. Ghofrani, "A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting," *Energy Conversion and Management*, vol. 118, pp. 331–344, 2016.

[26] A. Matteri, E. Ogliari, and A. Nespoli, "Enhanced Day-Ahead PV Power Forecast : Dataset Clustering for an Effective Artificial Neural Network Training †," vol. 5, no. 16, 2021.

[27] M.-F. Li, X.-P. Tang, W. Wu, and H.-B. Liu, "General models for estimating daily global solar radiation for different solar radiation zones in mainland china," *Energy Conversion and Management*, vol. 70, pp. 139–148, 2013.

[28] T. Ayodele, A. Ogunjuyigbe, A. Amedu, and J. Munda, "Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms," *Renewable Energy Focus*, vol. 29, pp. 78–93, 2019.

[29] P. F. Jiménez-Pérez and L. Mora-López, "Modeling and forecasting hourly global solar radiation using clustering and classification techniques," *Solar Energy*, vol. 135, pp. 682–691, 2016.

[30] Q. Cao, Y. Liu, K. Lyu, Y. Yu, D. H. Li, and L. Yang, "Solar radiation zoning and daily global radiation models for regions with only surface meteorological measurements in china," *Energy Conversion and Management*, vol. 225, p. 113447, 2020.

[31] J. Wu and C. K. Chan, "Prediction of hourly solar radiation with multi-model framework," *Energy Conversion and Management*, vol. 76, pp. 347–355, 2013.

[32] B. O. H. Ali-ou salah and T. Mouhaydine, "Short-term solar radiation forecasting using a new seasonal clustering technique and artificial neural network," *International Journal of Green Energy*, vol. 19, pp. 424–434, 2022.

[33] V. Vapnik, "Support-Vector Networks," *Machine Leaming*, vol. 20, no. 3, pp. 273–297, 1995.

[34] M. Rana, I. Koprinska, and V. G. Agelidis, "2D-interval forecasts for solar power production," *Solar Energy*, vol. 122, pp. 191–203, 2015.

[35] M. A. Ramli, S. Twaha, and Y. A. Al-Turki, "Investigating the performance of support vector machine and artificial neural networks in predicting solar radiation on a tilted surface: Saudi Arabia case study," *Energy Conversion and Management*, vol. 105, pp. 442–452, 2015.

[36] K. Mohammadi, S. Shamshirband, C. W. Tong, K. A. Alam, and D. Petkovi??, "Potential of adaptive neuro-fuzzy system for prediction of daily global solar

radiation by day of the year," *Energy Conversion and Management*, vol. 93, pp. 406–413, 2015.

[37] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.

[38] A. Mellit, A. M. Pavan, and M. Benghanem, "Least squares support vector machine for short-term prediction of meteorological time series," *Theoretical and applied climatology*, vol. 111, no. 1, pp. 297–307, 2013.

[39] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, and D. Heinemann, "Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement , numerical weather prediction , and cloud motion data," vol. 135, pp. 197–208, 2016.

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

[41] Y. W. Liu, H. Feng, H. Y. Li, and L. L. Li, "An improved whale algorithm for support vector machine prediction of photovoltaic power generation," *Symmetry*, vol. 13, no. 2, pp. 1–26, 2021.

[42] K. S. Chen, K. P. Lin, J. X. Yan, and W. L. Hsieh, "Renewable power output forecasting using least-squares support vector regression and google data," *Sustainability (Switzerland)*, vol. 11, no. 11, 2019.

[43] B. Mohammadi and Z. Aghashariatmadari, "Estimation of solar radiation using neighboring stations through hybrid support vector regression boosted by Krill Herd algorithm," *Arabian Journal of Geosciences*, vol. 13, no. 10, pp. 1–16, 2020.

[44] A. Rafati, M. Joorabian, E. Mashhour, and H. R. Shaker, "High Dimensional Very Short-Term Solar Power Forecasting Based on a Data-Driven Heuristic Method," *Energy*, vol. 7, no. 1, p. 119647, 2020.

[45] P. Anand, R. Rastogi, and S. Chandra, "A class of new Support Vector Regression models," *Applied Soft Computing Journal*, vol. 94, p. 106446, 2020.

[46] A. Danandeh Mehr, V. Nourani, V. Karimi Khosrowshahi, and M. A. Ghorbani, "A hybrid support vector regression–firefly model for monthly rainfall forecasting," *International Journal of Environmental Science and Technology*, vol. 16, no. 1, pp. 335–346, 2019.

[47] W. C. Hong and G. F. Fan, "Hybrid empirical mode decomposition with support vector regression model for short term load forecasting," *Energies*, vol. 12, no. 6, pp. 1–16, 2019.

[48] Z. Zhang and W.-c. Hong, "Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads," *Knowledge-Based Systems*, vol. 228, no. 1, p. 107297, 2021.

[49] A. K. Singh, Ibraheem, S. Khatoon, M. Muazzam, and D. K. Chaturvedi, "Load forecasting techniques and methodologies: A review," in *2012 2nd International Conference on Power, Control and Embedded Systems*, pp. 1–10, 2012.

[50] A. Hammer, D. Heinemann, and B. Luckehe, "Short-Term Forecasting of solar radaition : A statistical approach using satellite data," *Solar Energy*, vol. 67, no. 3, pp. 139–150, 1999.

[51] P. Blanc, J. Remund, and L. Vallance, "6 - short-term solar power forecasting based on satellite images," in *Renewable Energy Forecasting* (G. Kariniotakis, ed.), Woodhead Publishing Series in Energy, pp. 179–198, Woodhead Publishing, 2017.

[52] T. Strzy, "Day-Ahead Wind Power Forecasting in Poland Based on Numerical Weather Prediction," *Energies*, vol. 14, no. 0, p. 2164, 2021.

[53] C. Cornaro, M. Pierro, and F. Bucci, "Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast," *Solar Energy*, vol. 111, pp. 297–312, 2015.

[54] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881–2893, 2011.

[55] Z. Peng, D. Yu, D. Huang, J. Heiser, S. Yoo, and P. Kalb, "3D cloud detection and tracking system for solar forecast using multiple sky imagers," *Solar Energy*, vol. 118, pp. 496–519, 2015.

[56] R. Marquez, H. T. Pedro, and C. F. Coimbra, "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs," *Solar Energy*, vol. 92, pp. 176–188, 2013.

[57] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2012.

[58] A. a. Lacis and J. Hansen, "A Parameterization for the Absorption of Solar Radiation in the Earth's Atmosphere," 1974.

[59] P. E. Thornton and S. W. Running, "An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation," *Agricultural and Forest Meteorology*, vol. 93, pp. 211–228, 1999.

[60] Angstrom, "Angstrom.pdf," *Royal Metrological Society*, vol. 50, no. 210, pp. 121–126, 1924.

[61] A. Angstrom, "Computation of global radiation from records of sunshine," *Ark. Geofys.;(Sweden)*, vol. 2, pp. 471–479, Jan 1956.

[62] K. Sopian and M. Y. H. Othman, "Estimates of monthly average daily global solar radiation in Malaysia," *Renewable Energy*, vol. 2, no. 3, pp. 319–325, 1992.

[63] Z. Şen and E. Tan, "Simple models of solar radiation data for northwestern part of turkey," *Energy conversion and management*, vol. 42, no. 5, pp. 587–598, 2001.

[64] T. C. Chineke, "Equations for estimating global solar radiation in data sparse regions," *Renewable Energy*, vol. 33, no. 4, pp. 827–831, 2008.

[65] J. K. Yohanna, I. N. Itodo, and V. I. Umogbai, "A model for determining the global solar radiation for Makurdi, Nigeria," *Renewable Energy*, vol. 36, no. 7, pp. 1989–1992, 2011.

[66] T. Khatib, A. Mohamed, M. Mahmoud, and K. Sopian, "Modeling of daily solar energy on a horizontal surface for five main sites in Malaysia," *International Journal of Green Energy*, vol. 8, no. 8, pp. 795–819, 2011.

[67] Y. A. Abdalla and M. K. Baghdady, "Global and diffuse solar radiation in Doha (Qatar)," *Solar and Wind Technology*, vol. 2, no. 3-4, pp. 209–212, 1985.

[68] R. B. Benson, M. V. Paris, J. E. Sherry, and C. G. Justus, "Estimation of daily and monthly direct, diffuse and global solar radiation from sunshine duration measurements," *Solar Energy*, vol. 32, no. 4, pp. 523–535, 1984.

[69] M. Collares-Pereira and A. Rabl, "The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values," *Solar Energy*, vol. 22, no. 2, pp. 155–164, 1979.

[70] H. Li, W. Ma, X. Wang, and Y. Lian, "Estimating monthly average daily diffuse solar radiation with multiple predictors: A case study," *Renewable Energy*, vol. 36, no. 7, pp. 1944–1948, 2011.

[71] S. Topçu, S. DIlmaç, and Z. Aslan, "Study of hourly solar radiation data in Istanbul," *Renewable Energy*, vol. 6, no. 2, pp. 171–174, 1995.

[72] A. Trabea, "Analysis of solar radiation measurements at al-arish area, north sinai, egypt," *Renewable Energy*, vol. 20, no. 1, pp. 109–125, 2000.

[73] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[74] K. Bakirci, "Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey," *Energy*, vol. 34, no. 4, pp. 485–501, 2009.

[75] R. Huang, T. Huang, R. Gadh, and N. Li, "Solar generation prediction using the ARMA model in a laboratory-level micro-grid," *2012 IEEE 3rd International*

Conference on Smart Grid Communications, SmartGridComm 2012, pp. 528–533, 2012.

[76] G. Mihalakakou, M. Santamouris, and D. N. Asimakopoulos, "The total solar radiation time series simulation in Athens, using neural networks," *Theoretical and Applied Climatology*, vol. 66, no. 3-4, pp. 185–197, 2000.

[77] S. A. Kalogirou, "Artificial neural networks in renewable energy systems applications: a review," *Renewable & Sustainable Energy Reviews*, vol. 5, no. 4, pp. 373–401, 2001.

[78] K. S. Reddy and M. Ranjan, "Solar resource estimation using artificial neural networks and comparison with other correlation models," *Energy Conversion and Management*, vol. 44, no. 15, pp. 2519–2530, 2003.

[79] A. Mellit, M. Menghanem, and M. Bendekhis, "Artificial neural network model for prediction solar radiation data: application for sizing stand-alone photovoltaic power system," *IEEE Power Engineering Society General Meeting 2005*, no. 2, pp. 40–44 Vol. 1, 2005.

[80] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy*, vol. 84, no. 5, pp. 807–821, 2010.

[81] T. Khatib, A. Mohamed, K. Sopian, and M. Mahmoud, "Solar energy prediction for Malaysia using artificial neural networks," *International Journal of Photoenergy*, vol. 2012, 2012.

[82] S. Aggarwal and L. Saini, "Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest," *Energy*, vol. 78, pp. 247–256, 2014.

[83] S. Bhardwaj, V. Sharma, S. Srivastava, O. Sastry, B. Bandyopadhyay, S. Chandel, and J. Gupta, "Estimation of solar radiation using a combination of hidden markov model and generalized fuzzy model," *Solar Energy*, vol. 93, pp. 43–54, 2013.

[84] L. Suganthi, S. Iniyan, and A. A. Samuel, "Applications of fuzzy logic in renewable energy systems–a review," *Renewable and sustainable energy reviews*, vol. 48, pp. 585–607, 2015.

[85] M. Rizwan, M. Jamil, S. Kirmani, and D. Kothari, "Fuzzy logic based modeling and estimation of global solar energy using meteorological parameters," *Energy*, vol. 70, pp. 685–691, 2014.

[86] S. Jafarzadeh, M. S. Fadali, and S. Member, "Solar power prediction using interval type-2 TSK modeling," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 333–339, 2013.

[87] A. Mellit, A. H. Arab, N. Khorissi, and H. Salhi, "An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature," *2007 IEEE Power Engineering Society General Meeting, PES*, 2007.

[88] T. R. Sumithira and a. Nirmal Kumar, "Prediction of monthly global solar radiation using adaptive neuro fuzzy inference system (ANFIS) technique over the State of Tamilnadu (India): a comparative study," *Applied Solar Energy*, vol. 48, no. 2, pp. 140–145, 2012.

[89] W. Yaïci and E. Entchev, "Adaptive neuro-fuzzy inference system modelling for performance prediction of solar thermal energy system," *Renewable Energy*, vol. 86, pp. 302–315, 2016.

[90] S. Hussain and A. Al-Alili, "A new approach for model validation in solar radiation using wavelet, phase and frequency coherence analysis," *Applied Energy*, vol. 164, pp. 639–649, 2016.

[91] E. Akarslan and F. O. Hocaoglu, "A novel adaptive approach for hourly solar radiation forecasting," *Renewable Energy*, vol. 87, pp. 628–633, 2016.

[92] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[93] S. Belaid and A. Mellit, "Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate," *Energy Conversion and Management*, vol. 118, pp. 105–118, 2016.

[94] P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi, "A benchmarking of machine learning techniques for solar radiation forecasting in an insular context," *Solar Energy*, vol. 112, pp. 446–457, 2015.

[95] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine Learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[96] L. Olatomiwa, S. Mekhilef, S. Shamshirband, K. Mohammadi, D. Petković, and C. Sudheer, "A support vector machine–firefly algorithm-based model for global solar radiation prediction," *Solar Energy*, vol. 115, pp. 632–644, may 2015.

[97] S. Ghosh and R. Yadav, "Future of photovoltaic technologies: A comprehensive review," *Sustainable Energy Technologies and Assessments*, vol. 47, no. April, p. 101410, 2021.

[98] Z. Kherici, N. Kahoul, H. Cheghib, M. Younes, and B. Chekal Affari, "Main degradation mechanisms of silicon solar cells in Algerian desert climates," *Solar Energy*, vol. 224, no. February, pp. 279–284, 2021.

[99] J. M. R. S. P. P. Kim, "A Review of the Degradation of Photovoltaic Modules for Life Expectancy," *Energies*, vol. 14, no. 4, p. 4278, 2021.

[100] V. Dimitrievska, F. Pittino, W. Muehleisen, N. Diewald, M. Hilweg, A. Montvay, and C. Hirschl, "Statistical methods for degradation estimation and anomaly detection in photovoltaic plants," *Sensors*, vol. 21, no. 11, pp. 1–25, 2021.

[101] S. Chawla and V. A. Tikkiwal, "Performance evaluation and degradation analysis of different photovoltaic technologies under arid conditions," *International Journal of Energy Research*, vol. 45, no. 1, pp. 786–798, 2021.

[102] S. Djordjevic, D. Parlevliet, and P. Jennings, "Detectable faults on recently installed solar modules in Western Australia," *Renewable Energy*, vol. 67, pp. 215–221, 2014.

[103] M. Köntges, S. Kurtz, C. Packard, U. Jahn, K. Berger, K. Kato, T. Friesen, H. Liu, and M. Van Iseghem, "Performance and reliability of photovoltaic sys-

tems, subtask 3.2: review of failures of photovoltaic modules," *Report2014*, 2014.

[104] C. Dechthummarong, B. Wiengmoon, D. Chenvidhya, C. Jivacate, and K. Kirtikara, "Physical deterioration of encapsulation and electrical insulation properties of PV modules after long-term operation in Thailand," *Solar Energy Materials and Solar Cells*, vol. 94, no. 9, pp. 1437–1440, 2010.

[105] T. Kojima and T. Yanagisawa, "The evaluation of accelerated test for degradation a stacked a-Si solar cell and EVA films," *Solar Energy Materials and Solar Cells*, vol. 81, no. 1, pp. 119–123, 2004.

[106] G. Oreski and G. M. Wallner, "Evaluation of the aging behavior of ethylene copolymer films for solar applications under accelerated weathering conditions," *Solar Energy*, vol. 83, no. 7, pp. 1040–1047, 2009.

[107] N. Mishra, A. S. Yadav, R. Pachauri, Y. K. Chauhan, and V. K. Yadav, "Performance enhancement of pv system using proposed array topologies under various shadow patterns," *Solar Energy*, vol. 157, pp. 641–656, 2017.

[108] J. H. Wohlgemuth and S. Kurtz, "Reliability testing beyond qualification as a key component in photovoltaic's progress toward grid parity," *IEEE International Reliability Physics Symposium Proceedings*, pp. 551–556, 2011.

[109] M. A. Quintana, D. L. King, T. J. McMahon, and C. R. Osterwald, "Commonly observed degradation in field-aged photovoltaic modules," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 1436–1439, 2002.

[110] D. E. Carlson, R. Romero, F. Willing, D. Meakin, L. Gonzalez, R. Murphy, H. R. Moutinho, and M. Al-Jassim, "Corrosion Effects in Thin-Film Photovoltaic Modules," *Progress in Photovoltaics: Research and Applications*, vol. 11, no. 6, pp. 377–386, 2003.

[111] M. D. Kempe, "Control of moisture ingress into photovoltaic modules," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 503–506, 2005.

[112] M. A. Munoz, M. C. Alonso-García, N. Vela, and F. Chenlo, "Early degradation of silicon PV modules and guaranty conditions," *Solar Energy*, vol. 85, no. 9, pp. 2264–2274, 2011.

[113] S. Chattopadhyay, R. Dubey, V. Kuthanazhi, J. J. John, C. S. Solanki, A. Kottantharayil, B. M. Arora, K. Narasimhan, J. Vasi, B. Bora, *et al.*, "All india survey of photovoltaic module degradation 2014: survey methodology and statistics," in *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*, pp. 1–6, IEEE, 2015.

[114] S. Chattopadhyay, R. Dubey, V. Kuthanazhi, J. J. John, C. S. Solanki, A. Kottantharayil, B. M. Arora, K. L. Narasimhan, V. Kuber, J. Vasi, A. Kumar, and O. S. Sastry, "Visual degradation in field-aged crystalline silicon PV modules in India and correlation with electrical degradation," *IEEE Journal of Photovoltaics*, vol. 4, no. 6, pp. 1470–1476, 2014.

[115] M. Piliougine, A. Oukaja, P. Sánchez-Friera, G. Petrone, F. J. Sánchez-Pacheco, G. Spagnuolo, and M. Sidrach-de Cardona, "Analysis of the degradation of single-crystalline silicon modules after 21 years of operation," *Progress in Photovoltaics: Research and Applications*, vol. 29, no. 8, pp. 907–919, 2021.

[116] E. Kaplani, "Degradation in field-aged crystalline silicon photovoltaic modules and diagnosis using electroluminescence imaging," *8th International Workshop on Teaching in Photovoltaics*, no. April, pp. 38–41, 2016.

[117] E. Kaplani, *PV Cell and Module Degradation, Detection and Diagnostics*, pp. 393–402. Cham: Springer International Publishing, 2016.

[118] A. Kyprianou, A. Phinikarides, G. Makrides, and G. E. Georghiou, "Robust Principal Component Analysis For Computing The Degradation Rates Of Different Photovoltaic Systems," *29th EU-PVSEC*, vol. 58, no. 3, pp. 2939–2942, 2014.

[119] D. C. Jordan, B. Sekulic, B. Marion, and S. R. Kurtz, "Performance and aging of a 20-year-old silicon pv system," *IEEE Journal of Photovoltaics*, vol. 5, no. 3, pp. 744–751, 2015.

[120] P. Rajput, G. Tiwari, O. Sastry, B. Bora, and V. Sharma, "Degradation of mono-crystalline photovoltaic modules after 22years of outdoor exposure in the composite climate of india," *Solar Energy*, vol. 135, pp. 786–795, 2016.

[121] P. Rajput, O. Sastry, and G. Tiwari, "Effect of irradiance, temperature exposure and an arrhenius approach to estimating weathering acceleration factor of glass, eva and tedlar in a composite climate of india," *Solar Energy*, vol. 144, pp. 267–277, 2017.

[122] P. Rajput, G. Tiwari, and O. Sastry, "Thermal modelling and experimental validation of hot spot in crystalline silicon photovoltaic modules for real outdoor condition," *Solar Energy*, vol. 139, pp. 569–580, 2016.

[123] A. Phinikarides, N. Kindyni, G. Makrides, and G. E. Georghiou, "Review of photovoltaic degradation rate methodologies," *Renewable and Sustainable Energy Reviews*, vol. 40, pp. 143–152, 2014.

[124] A. Livera, A. Phinikarides, G. Makrides, and G. E. Georghiou, "Impact of missing data on the estimation of photovoltaic system degradation rate," in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, pp. 1954–1958, IEEE, 2017.

[125] L. Tan, K. Yu, F. Ming, X. Cheng, and G. Srivastava, "Secure and resilient artificial intelligence of things: A honeynet approach for threat detection and situational awareness," *IEEE Consumer Electronics Magazine*, vol. 11, no. 3, pp. 69–78, 2022.

[126] J. Y. Ye, T. Reindl, A. G. Aberle, and T. M. Walsh, "Performance degradation of various pv module technologies in tropical singapore," *IEEE Journal of Photovoltaics*, vol. 4, no. 5, pp. 1288–1294, 2014.

[127] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust pv degradation methodology and application," *IEEE Journal of photovoltaics*, vol. 8, no. 2, pp. 525–531, 2017.

[128] D. C. Jordan, M. G. Deceglie, and S. R. Kurtz, "Pv degradation methodology comparison—a basis for a standard," in *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, pp. 0273–0278, IEEE, 2016.