REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

FORECASTING OF SOLAR RADIATION FOR SOLAR SYSTEM UNDER DIFFERENT CLIMATIC CONDITION USING SATELLITE IMAGERY

Khairi Sameer Khairi KHAIRI

Master's Thesis

Supervisor

Assoc. Prof. Dr. Adil DENIZ DURU

Istanbul, 2022

FORECASTING OF SOLAR RADIATION FOR SOLAR SYSTEM UNDER DIFFERENT CLIMATIC CONDITION USING SATELLITE IMAGERY

Khairi Sameer Khairi KHAİRİ

Electrical and Computer Engineering

Master’s Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled FORECASTING OF SOLAR RADIATION FOR SOLAR SYSTEM UNDER DIFFERENT CLIMATIC CONDITION USING SATELLITE IMAGERY prepared by KHAIRI SAMEER KHAIRI KHAIRI and submitted on 22/12/2022 has been **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering.

Assoc. Prof. Dr. Adil Deniz DURU

the Supervisor

Thesis Defense Committee Members:

Prof. Dr. Oguz BAYAT	Department of Software Engineering, Altınbaş University	_____
Asst. Prof. Dr. Adullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	_____
Assoc. Prof. Dr. Adil Deniz DURU	Department of Sports and Health Sciences, Marmara University	_____

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: ___/___/___

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Khairi Sameer Khairi KHAIRI

Signature

ABSTRACT

FORECASTING OF SOLAR RADIATION FOR SOLAR SYSTEM UNDER DIFFERENT CLIMATIC CONDITION USING SATELLITE IMAGERY

Khairi, Khairi Sameer Khairi

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Assoc. Prof. Dr. Adil DENIZ DURU

Date: December / 2022

Pages: 74

As it is commonly used to predict power output, accurate solar irradiance forecasting is critical for minimizing operational costs of solar photovoltaic (PV) generation. The goal of solar irradiance prediction is to aid in the management of the electric grid, among other things. This is a growing requirement as solar energy grows in popularity. Sun oriented irradiance is the power gotten from the Sun as electromagnetic radiation per unit region. Many external factors such as physical and external factors can affect how much solar irradiance is available at a given location on earth. The effective irradiance is called Global Horizontal Irradiance (GHI). In this paper, we propose a deep learning approach to solar irradiance nowcasting. The experimental results show that the proposed method achieves forecast skill of about 110 GHI RMSE for up to six hours in the future.

Keywords: Solar Irradiance, Machine Learning, Power Plant.

TABLE OF CONTENTS

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

GHI	:	Global Horizontal Irradiance
MAE	:	Mean Absolute Error
MSE	:	Mean Squared Error
NaN	:	Not a Number
PV	:	photovoltaic
RMSE	:	Root Mean Squared Error

1. INTRODUCTION

Renewable Energy is turning into an innovation and a perpetually feasible option in contrast to customary non-environmentally friendly power sources. Subsequently, the tenacious risk of environmental change powers humankind to take a stab at new and more successful ways of creating energy. Especially when we base all our essential necessities on this type of energy. As per the Worldwide Energy Organization (IEA) [1], worldwide inexhaustible power limit is projected to ascend by more than 1 TW, a 46 percent increment over the period 2018 to 2023. Sun based photovoltaic (PV) addresses the greater part of this development and overwhelms the development of sustainable capacity. However, because the energy output of PV panels depends on weather conditions such as cloud cover and solar irradiance, the PV panels' energy output is unstable. To understand and manage the output variability is of interest for several actors in the energy market. Nowadays, the use of Artificial Intelligence Technologies is becoming increasingly widespread due to their ability to solve highly complex problems. The expansion in the productivity of PCs and calculations has assisted with tackling issues, in designing as well as in many fields, for example, medicine [2], finance [3], and the environment [4]. For that reason, these instruments are turning out to be more well-known and can in any case gain some headway and can be executed in additional areas or issues. Machine learning is a computer science subfield, and it is categorized as a form of artificial intelligence. It can be used in many domains and the advantage of this approach is that a model can solve problems that clear algorithms cannot represent. These AI models are used in the occasion of Sustainable power Source (RES) to figure the energy created in a power station or even the activities of weather patterns. Such gauges are critical in the energy market. The issue of sun-based irradiance, changeability and unusualness that arrives at the World's surface is commonly known. An exact gauge of this variable will consequently empower better preparation and activity of force conveyance at the monetary or energy yield levels, either by making elective plans for customary power and in general schedules or by money management the suitable measure of energy assets and stores to limit the influence framework's working expenses.

Additionally, related literature on the prediction of solar surface irradiation, estimation of PV capacity and current state-of-the-art studies is reviewed [5]. Machine learning algorithms such as support vector machine (SVM) [6,7], k-nearest neighbor [8,9], and Random Forest [10] have been used in studies; others compared or combined the prediction results with multiple machine learning models. Deep learning is a particular subfield of machine learning designed to allow machines to simulate the way the human brain thinks, and its operational model is based on neuroscience [11]. Deep learning is designed for the representation of input and target data using a neural network structure. Actually, several types of research have addressed solar irradiance prediction based on deep learning models. In [12], the creators have proposed a profound learning half breed model to foresee sun-based radiation in 2 phases, a convolutional network is utilized to remove highlights, and a Long-Transient Memory (LSTM) network is utilized for the expectation stage. [13] proposed an intermittent brain network model to explore how arising profound learning calculations, particularly contrasted with gulping fake brain organizations, add to precise sun-based radiation forecast. Guijo-Rubio et al. [14] have introduced how transformative fake brain network calculations can acquire incredibly exact sun-oriented radiation forecast, solely working with satellite-based information. To predict solar irradiation from weather measurements, an advanced machine learning-based model was used in [15]. This model is based on LSTM networks which, in time series data, have an extended capacity to explain time dependencies. A deep learning approach based on traditional neural networks has been used to predict solar irradiance at thirty stations in Turkey [16].

1.1 MOTIVATION

Because of concerns about greenhouse gas emissions and environmental pollution caused by excessive consumption of fossil fuel energy sources, renewable energy sources have become increasingly important in power generation [1]. Besides, petroleum product energy supplies are restricted and will ultimately run out. Therefore, elective energy sources should be tried to supplant petroleum products. Toward the finish of 2018, power age from new wind and photovoltaic (PV) plants has become more practical than power age from petroleum product terminated plants in certain areas [28]. Besides, in certain areas, putting in new wind and sun-

oriented PV plants costs not exactly proceeding to work existing non-renewable energy source power plants [2]. Because solar energy is a clean, abundant, free, and sustainable source of energy, the deployment of PV panels to generate electricity from solar energy has expanded in recent years [3]. PV panel installation is projected to rise in the future due to governments' interest in investing in renewable energy sources. Figure 1.1 depicts global solar photovoltaic capacity from 2008 to 2018 [33][34][35].

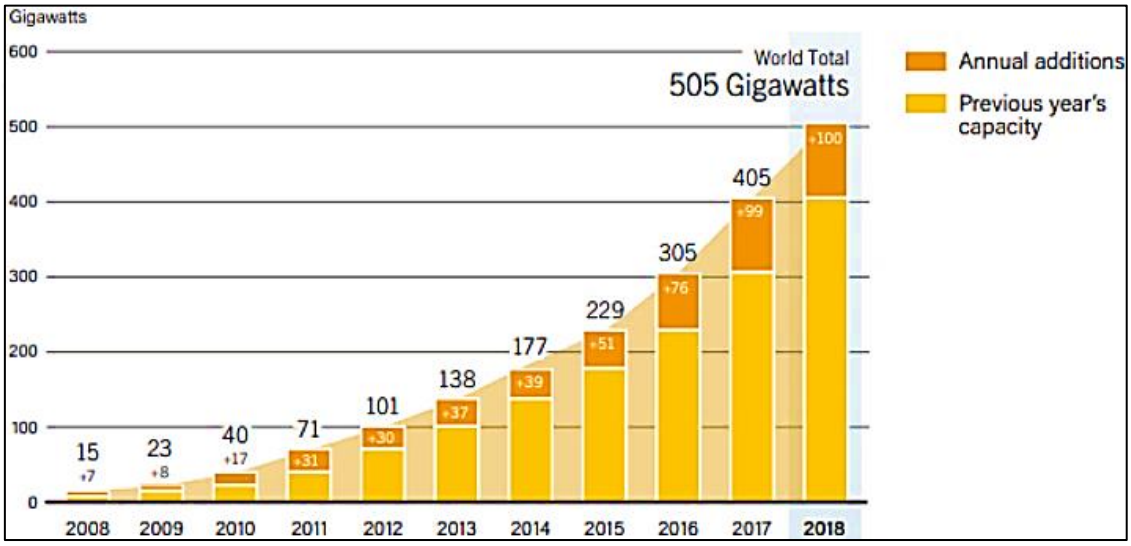

Figure 1.1: Solar PV global capacity and annual additions [2]

Since sun-oriented photovoltaics is one of the quickest creating energy advancements, and Italy is one of the best ten nations as far as sun-based PV limit, research in these disciplines is basic and undeniable. Figure 1.2 portrays the main ten nations with the biggest sunlight-based PV limit around the world [38][2].

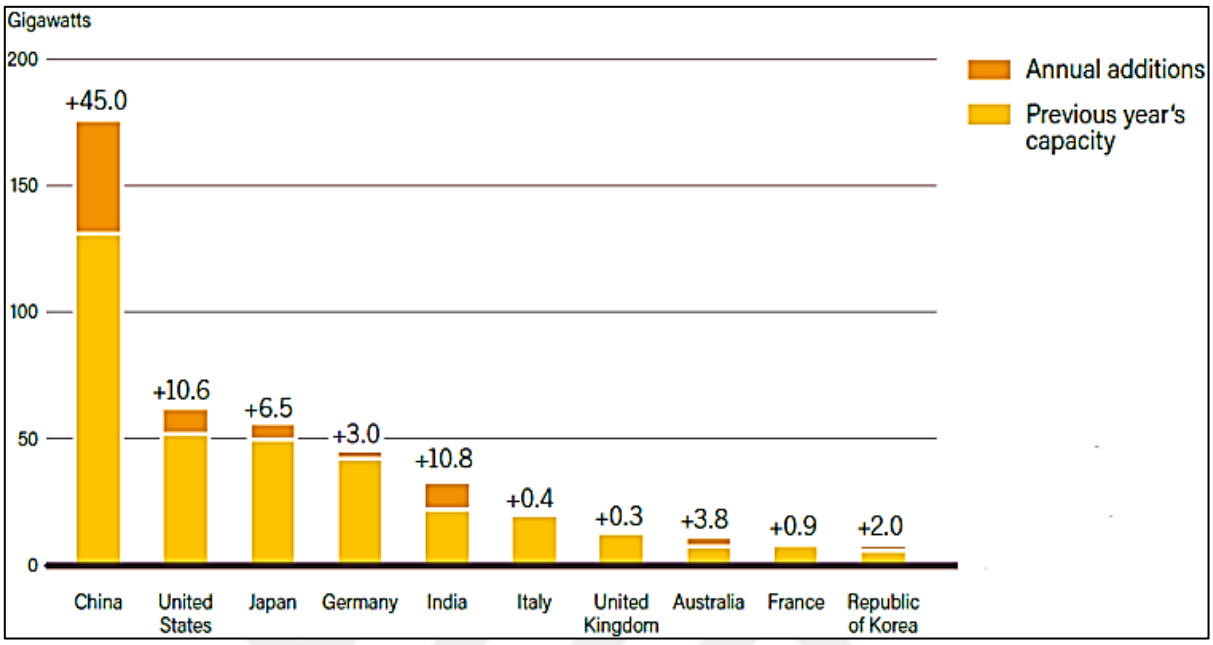

Figure 1.2: Solar PV capacity of top 10 countries [33]

Photovoltaic panels are increasingly being used in commercial and residential applications. Integrating renewable resources in microgrids is becoming increasingly popular as technology and economic viability improve. Microgrids are limited energy networks that create power utilizing nearby energy sources, for example, sunlight-based energy and wind power. A microgrid is an assortment of burdens, power age units, and energy stockpiling gadgets that cooperate [39][40].

Sun oriented Power is a notable Environmentally friendly power Source (RES) with gigantic potential. Nonetheless, on the grounds that it is subject to outside atmospheric conditions, this innovation is fairly factor and some of the time unwanted concerning both energy and financial matters. Because of their ability to solve highly complex problems, Artificial Intelligence Tools are becoming increasingly popular. The advancement of computer and algorithm performance aided in the solution of problems not only in engineering, but also in many other fields such as medicine, finance, and literature. Therefore, these apparatuses are turning out to be more well known, and they can in any case gain ground and be applied to

additional areas or issues. In Table 1.1 we present a few instances of use of Man-made reasoning Devices in various review fields.

Table 1.1: Fields of Application of Artificial Intelligence Tools

Field	Description
Automotive	Automotive automatic guidance systems, automatic braking systems, misfire detection, virtual emission sensors
Banking	Credit application evaluators, cash forecasting, firm classification, exchange rate forecasting
Electronics	Code sequence prediction, process control, chip failure analysis, nonlinear modeling
Medical	Breast cancer cell analysis, EEG and ECG analysis, optimization of transplant times, emergency room test advisement
Speech	Speech recognition, speech compression, vowel classification, text to speech synthesis
Telecommunications	Image and data compression, automated information services, real-time translation of spoken language, customer payment processing systems

On account of RES, these man-made reasoning devices are utilized to estimate how much energy delivered by a power station as well as the way of behaving of weather patterns. These are key figures in the Energy Market. The fluctuation and flightiness of sunlight-based illumination arriving at the World's surface are notable issues. Consequently, an exact figure of this variable can move a predominant Power Spread status and activity at the money related or energy creation level, either by making substitute game plans for typical impact and general timetables or committing the best degree of energy assets and stores to reduce the impact structure utilitarian expenses. Beside these advantages, a decent figure can assist with decreasing the effect of PV yield vulnerability on the framework, further develop framework dependability, keep up with power quality, and increment PV framework entrance. These devices' conjectures can likewise be utilized to explore the feasibility of sunlight-based power projects in a particular area.

There were a few convincing purposes behind choosing this proposition subject. The assessment of the meaning of this sort of advancement, as well as the potential that it can bring to the world's necessities, understanding of the ongoing circumstance in the Energy Area, the essential qualities of the resource used in this kind of development, and the crucial benefits of expecting Daylight based Irradiance became key components in the subject framework.

1.2 THESIS OBJECTIVES

Because of the significance of a decent Sun oriented Figure referred to previously, the fundamental goal of this proposal is to grasp what strategy, among the concentrated-on techniques, presents the best presentation in Sun powered Guaging. That's what to do, first we will examine on the accessible Man-made consciousness procedures and their applications on Sun oriented Irradiance Conjecture. Then, we will carry out a Sun oriented Irradiance Gauge model utilizing python Language and the procedures concentrated previously. To accomplish the most ideal correlation, the two methodologies should be enough matched to the current circumstance. Thus, to defeat a few imperfections and issues inborn in each model, undertaking a minor streamlining of the two strategies before looking at their performance is required.

2. LITERATURE SURVEY

This part endeavors to give a short outline of the examination embraced by different specialists in the subject of study. It gives a concise survey of the different methodologies and their qualities.

2.1 SOLAR ENERGY

Sun powered energy has been utilized as a wellspring of light and intensity since old times. The eighteenth century denoted the start of another time in sunlight-based energy innovation. In 1767, Horace de Saussure made the primary sun-situated finder, a compartment encompassed by three layers of glass to protect atomic power. In 1839, Edmond Becquerel found photovoltaic impact, which changes over daylight into electric stream in a substance [5]. In 1876, William Grylls Adams and Richard Evans Day showed the way that a solid substance can change over light into power without the utilization of extra warming or complex parts. Clarence Kemp imagined the primary business sunlight-based water warmer in 1891. William J. Bailley concocted a sun powered gatherer made of copper loops and a protected box in 1908. His revelation is the establishment for the present hardware. Sun oriented controlled hardware turned out to be more well known in the US after WWII [6]. Sun based energy has been utilized in space apparatus starting around 1958. During the 1970s, sunlight-based cell proficiency and cost were moved along. Numerous legislatures, including the US, have laid out sunlight-based energy research labs and establishments [5]. The main sunlight based fueled airplane and vehicle were worked in the mid-1980s [7]. Enormous scope sunlight-based power plants were planned and worked in the last part of the 1980s [7]. Sun based power plants were equipped for creating in excess of 200 Megawatts of force in 2013. Toward the finish of 2012, complete introduced sun-based power age limit had ascended to in excess of 100 GW, up from almost 2.2 GW in the mid-1990s [1]. Figure 2-1 portrays the worldwide absolute introduced PV power.

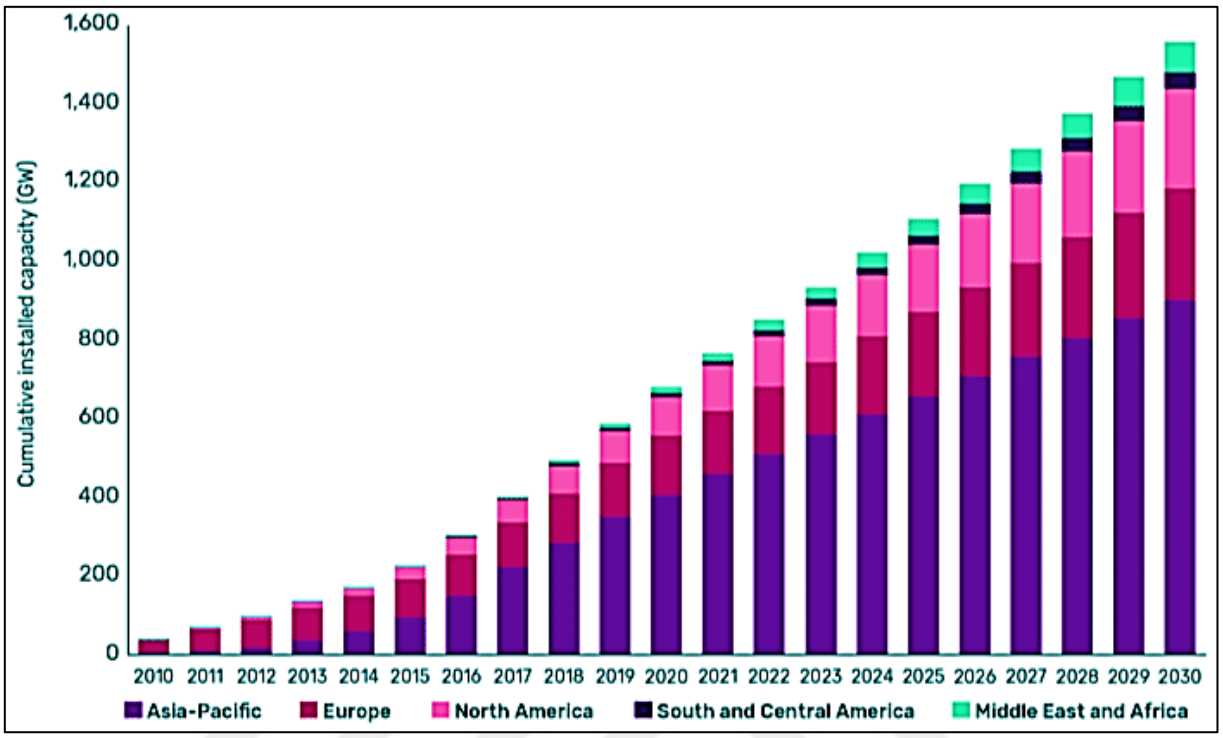

Figure 2.1: Total World Installed PV [1]

This chart shows that PV power establishments are expanding fundamentally for many years. It is likewise significant that the rising pattern of introducing PV frameworks started in 2006. Sun oriented energy can be reaped and changed over into power utilizing two distinct advancements. Sun based nuclear energy age and photovoltaic (PV) power age are two of these innovations.

2.1.1 Solar Thermal Technology

Sun based nuclear power innovation concentrates daylight to produce heat. This intensity can straightforwardly be utilized for rural drying, sun powered air warmers, sun based water radiators, sun oriented cooling frameworks, and sun oriented cookers. This intensity can likewise be utilized to control turbines on electrical generators. Figure 2.2 portrays an illustration of a sun based nuclear power age site [8].

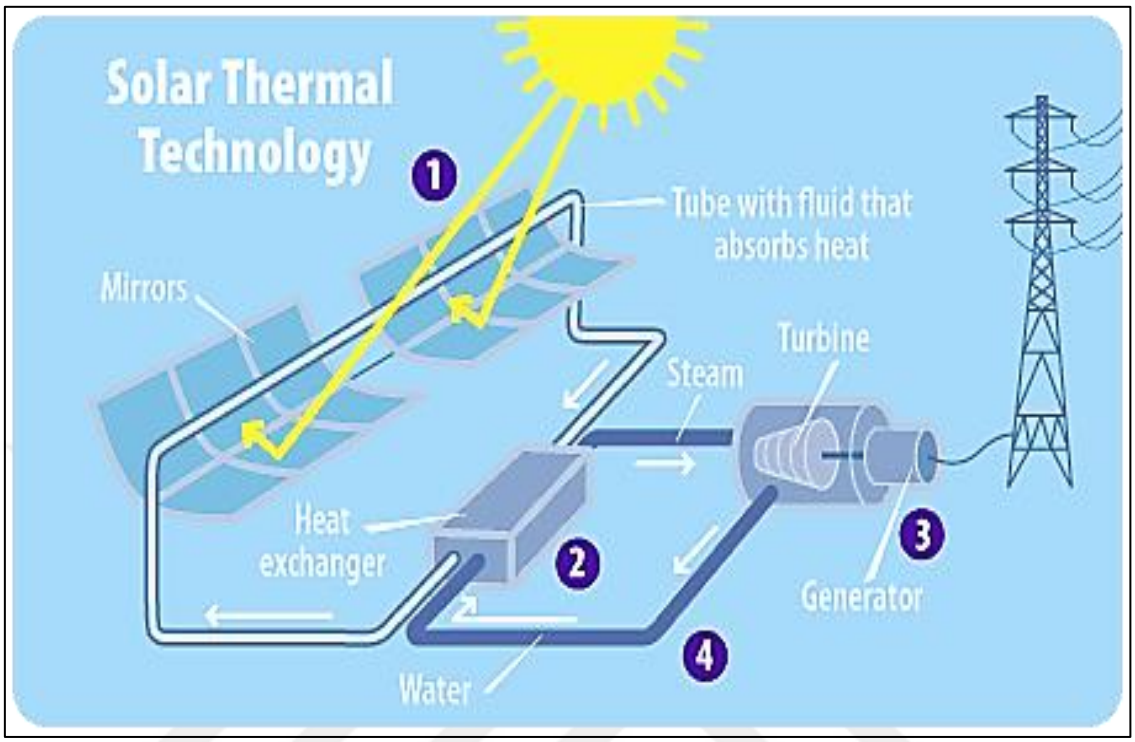

Figure 2.2: Solar Thermal Technology [8]

2.2 DISTRIBUTION OF SOLAR ENERGY

The electromagnetic radiation transmitted by the sun is known as sun-oriented radiation. The aggregate sum of quantum energy delivered by occurrence photons per unit region is known as sun powered radiation. Joules per square meter are units of estimation for sunlight-based radiation. This episode of energy esteem on the world's surface is impacted by variables like area, air contamination, and overcast cover [21].

Light photons connect with and enter the environment in four ways. The essential sort of cooperative exertion incorporates photons being reflected once more into space because of correspondence with the environment and Earth. Accordingly, the photons are eaten up by the worldwide climate. Third, photons disperse, bringing about diffuse radiation, which is significant in sun-controlled applications and choosing. At long last, the photons that are not consumed, dispersed, or reflected are alluded to as immediate radiation, which is the main figure sunlight-based power applications and estimating. Figure 2.3 portrays four cases of

photons entering the environment. Extraterrestrial sunlight-based energy misfortunes in the environment can arrive at 30% on a sunny morning and 100 percent on an overcast day [22, 23].

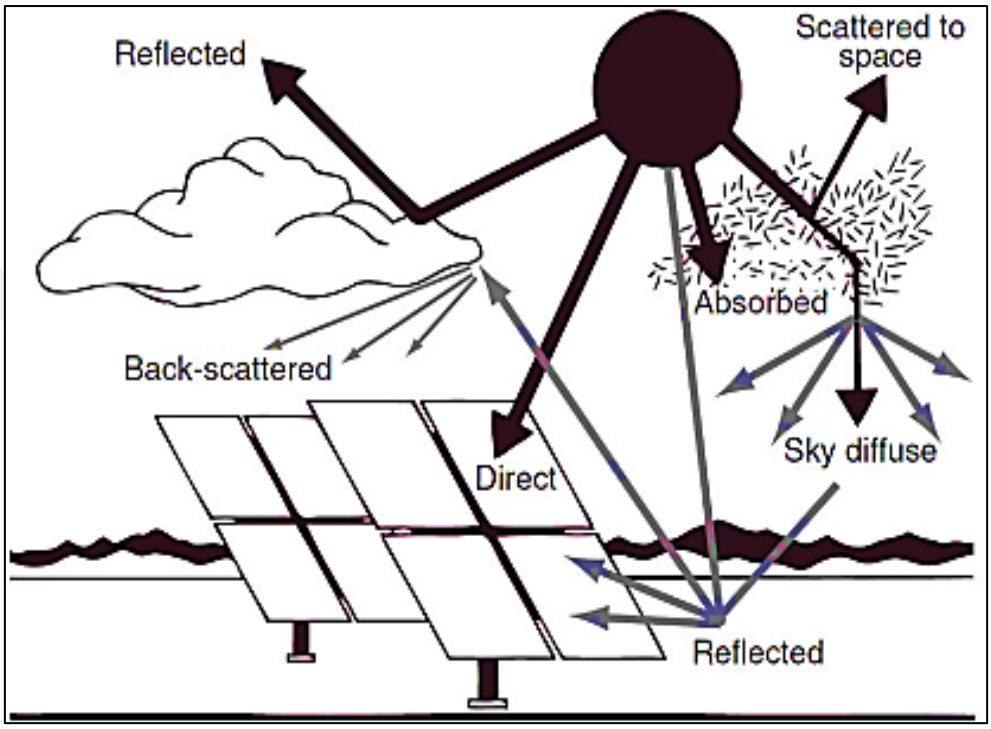

Figure 2.3: Solar Radiation Atmospheric Interaction [23]

Overall level irradiance (GHI) is the amount of direct sun-based radiation (Direct Normal Irradiance (DNI)) and diffuse level irradiance (GHI) that happens on an even surface (DHI). GHI can be composed as [24]:

$$GHI = DHI + DNI \cdot \cos(z) \tag{2.1}$$

where z is the solar zenith angle.

2.2.1 Solar Radiation in the World

The dissemination of sun-oriented radiation on Earth can be partitioned into three districts. Figure 2.4 portrays the worldwide conveyance of sun-based radiation. Significant differences in solar radiation are expected due to climate differences between regions of the world.

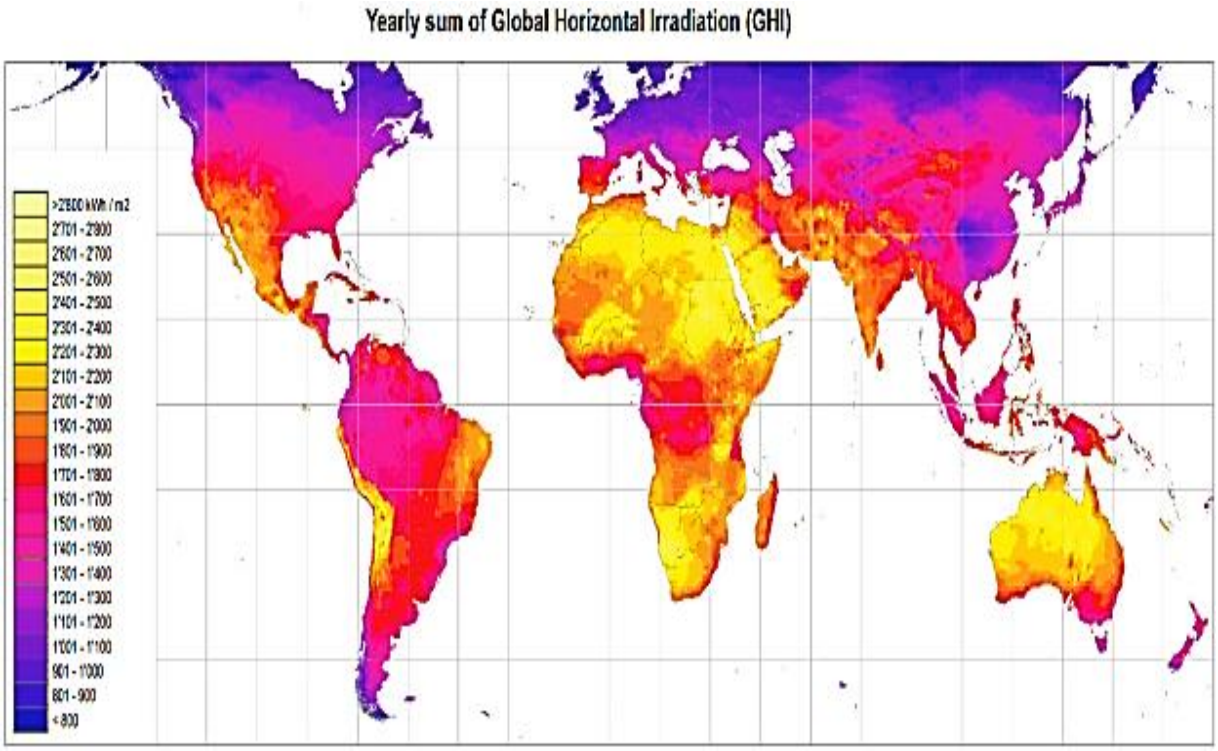

Figure 2.4: World Solar Radiation Intensity [25]

The main area gets the most sunlight-based radiation and is situated between scopes 15°N and 35°N. As a result of the absence of overcast cover and precipitation, most of this sunlight-based radiation is immediate radiation. The subsequent area is situated somewhere in the range of 15°N and 0°N scope. This area gets the second most sun based radiation. In view of the dampness and overcast cover, this locale has more significant levels of diffuse radiation. Over time, how much sun based radiation in this area is almost consistent. The third locale, above 35°N, has minimal measure of sun-based radiation. The overcast cover is at fault for this. Dispersion radiation represents almost 50% of the radiation around here. Sun powered radiation is at its most minimal at the World's posts [26].

2.3 SOLAR IRRADIANCE

The power got from the Sun as electromagnetic radiation per unit region is alluded to as sun-based irradiance. Watts per square meter (W/m2) is the unit of estimation for sun powered light. Solar irradiance research and measurement are important for predicting energy generation from solar power plants. There are several types of measured solar irradiance that are relevant to the problem discussed in this thesis:

i- Direct Ordinary Irradiance (DNI), the sun-based radiation estimated at an earth surface inverse to the sun, otherwise called bar irradiance. It just measures direct radiation from the sunlight based plate and doesn't consider diffuse radiation. [11].

ii- Diffuse Horizontal Irradiance (DHI) is the radiation estimated on Earth's even surface brought about by light dissipated by the climate. It estimates radiation from all places overhead with the exception of the sun's plate. There ought to be practically no diffuse sky radiation without even a trace of an environment. [11].

iii- The radiation reflected by non-climatic components, for example, the ground is known as reflected radiation. Notwithstanding, in light of the fact that sunlight powered chargers are commonly shifted away from the reflected radiation direction, it has minimal bearing on the complete radiation got by their surface. A special case is the point at which the surface is encircled by snow, which can essentially build how much reflected radiation got.

iv- The all out irradiance from the sun on the World's level surface is alluded to as worldwide flat irradiance (GHI). Not entirely set in stone by joining DHI, DNI (subsequent to adapting to the sun oriented peak point z) and reflected radiation. Since reflected radiation is many times unimportant in contrast with immediate and diffuse radiation, worldwide even radiation is characterized exclusively as the amount of immediate and diffuse radiation. [11]:

$$G = G_d + G_b cos(\theta_Z) \tag{2.3}$$

where G stands for Global Horizontal Irradiance, Gd stands for Diffuse Horizontal Irradiance, Gb stands for Direct Normal Irradiance or beam irradiance, and Z stands for Zenith Angle. Figure 2.5 depicts the various types of solar radiation mentioned.

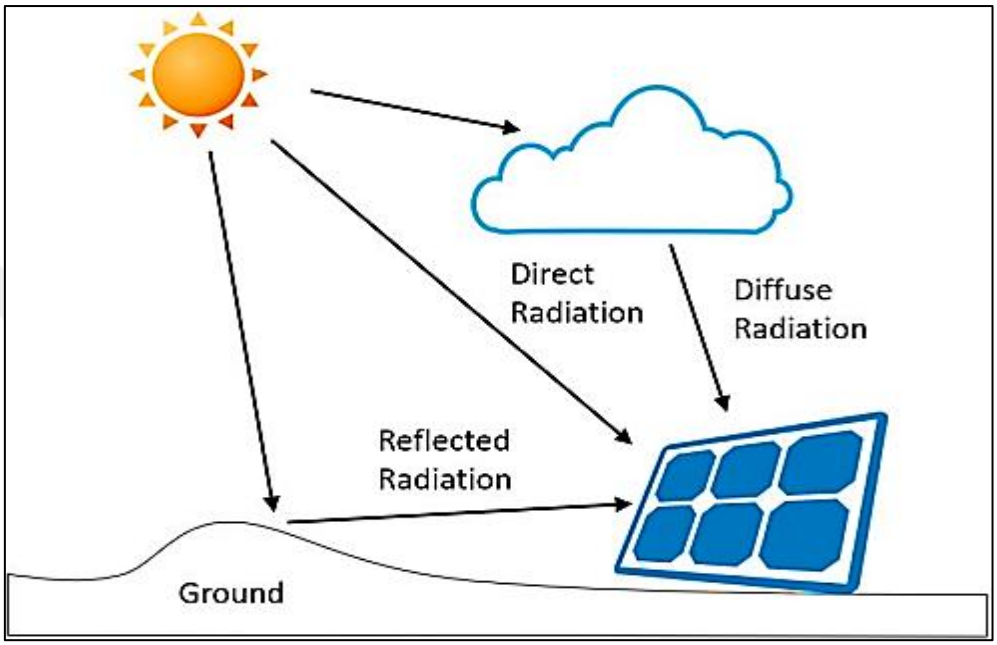

Figure 2.5: Types of solar radiation [11].

2.4 TYPE OF SOLAR RADIATION

2.4.1 Direct Normal Irradiance (DNI)

The photovoltaic radiation measured at a floor of the earth perpendicular to the Sun is also known as beam irradiance. It only measures direct solar disk radiation and excludes diffuse radiation [18].

2.4.2 Diffuse Horizontal Irradiance (DHI)

Is the radiation estimated on a flat Earth surface from gentle dissipated by the environment. With the exception of radiation from the sun-based plate, it estimates radiation from all places overhead. There should be basically no diffuse sky radiation without any an air [18].

2.4.3 Reflected Radiation

Is the radiation reflected by non-environmental factors like the ground? In any case, in light of the fact that photovoltaic boards are shifted away from the reflected radiation direction, it is seldom applicable in the all-out radiation procured by utilizing their surface. An exemption is the point at which the floor is encircled by snow, which expands the size of the reflected radiation got [18].

2.4.4 Global Horizontal Irradiance (GHI)

Is the total solar irradiance on Earth's horizontal floor. It is the sum of DHI, DNI (after accounting for the solar z's photovoltaic zenith perspective), and mirrored radiation. Be that as it may, on the grounds that reflected radiation is habitually irrelevant in contrast with immediate and diffuse radiation for every sensible capability, global level radiation is characterized exclusively as the amount of immediate and diffuse radiation:

$$G = G_d + G_b cos(\theta_z) \tag{2.4}$$

G addresses Worldwide Flat Irradiance, Gd addresses Diffuse Even Irradiance, Gb addresses Direct Typical Irradiance or pillar irradiance, and z addresses the pinnacle point. Figure 2.6 illustrates the various types of photovoltaic radiation:

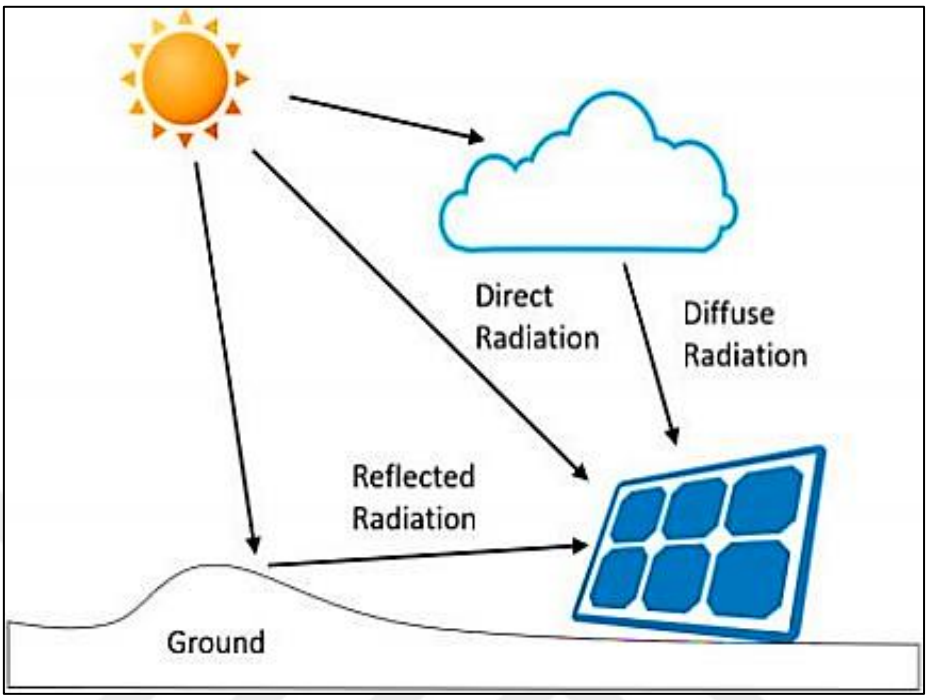

Figure 2.6: Types of Solar Radiation [18].

The energy, Eph , of each photon is directly related to the wavelength λ by the relationship

$$E_{ph} = \frac{hC}{\lambda} \tag{2.5}$$

Where h is the Planck constant, h = 6,62.10−34 J.s [18].

2.5 SOLAR PHOTOVOLTAIC ENERGY

Solar photovoltaic energy has been of great interest in recent years. It is non-polluting energy and provides real solutions to the various problems that The European Union and its Member States are currently discussing climate change and the energy crisis [20].

2.5.1 Photovoltaic Effect

The photovoltaic impact is an actual peculiarity that happens just specifically materials known as semiconductors (the most popular is silicon). It quickly changes energy from sunlight-based beams into power by delivering and moving positive or negative electrical burdens in a

semiconductor material because of the light. The electrical energy obtained is called photovoltaic energy [21].

2.5.2 Photovoltaic System

The photovoltaic framework is comprised of a field of modules and an assortment of parts that adjust the power delivered by the modules to receptor specifications [21]. The succinct outline of an independent photovoltaic framework is found in the image underneath.

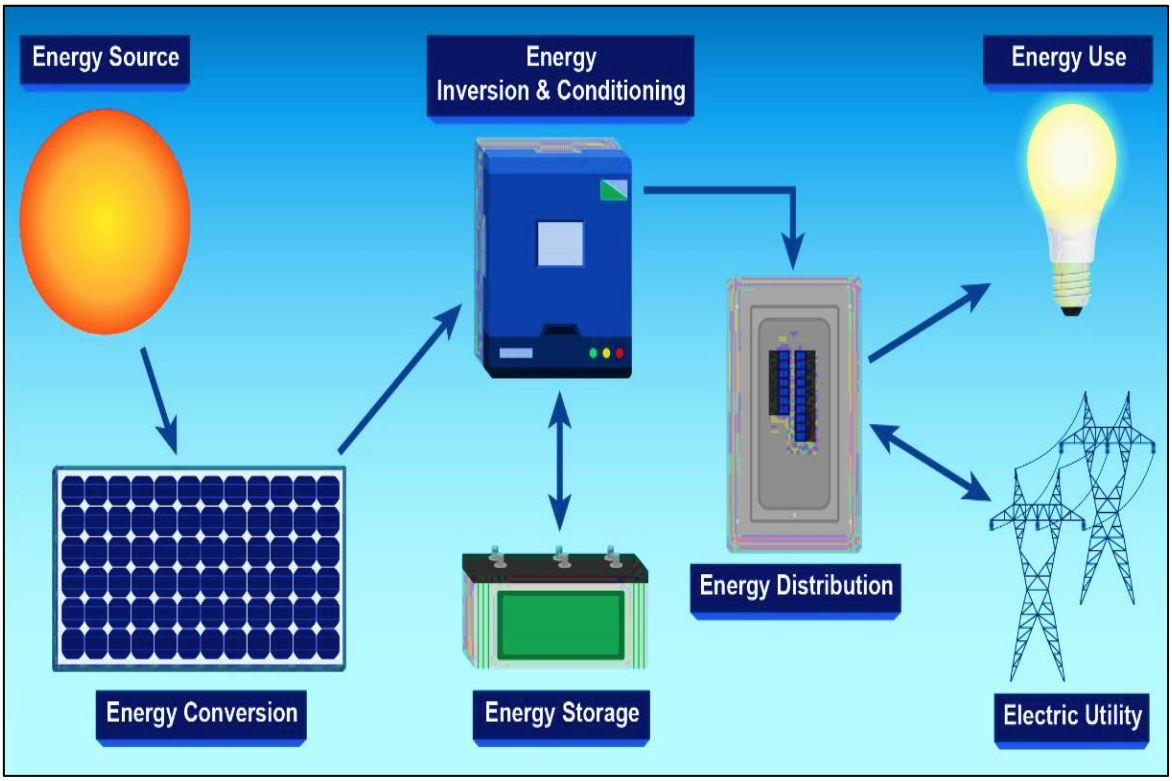

Figure 2.7: Diagram of an autonomous photovoltaic system [21].

2.5.3 Photovoltaic Cell

The photovoltaic cell is a thin layer (two-tenths of a millimeter) of a semiconductor metal, processed to be able to convert light into direct current. It consists of the stacking of two layers of silicon previously exposed to ion beams, one to phosphorus (-) ions and the other to boron (+) ions. The first layer has an excess of electrons and the other a deficit. They are said to be

respectively doped N and doped P. This process is called the «doping» and serves to create an electric field between the two zones where is created a junction called PN and directed from the zone (P) to the zone (N). The region (N) is covered by a metal framework that goes about as a cathode K while a metal plate A covers the opposite side of the gem and goes about as an anode. A light shaft that strikes the gadget can enter the gem through the matrix and influence a voltage to show up between the cathode and anode.

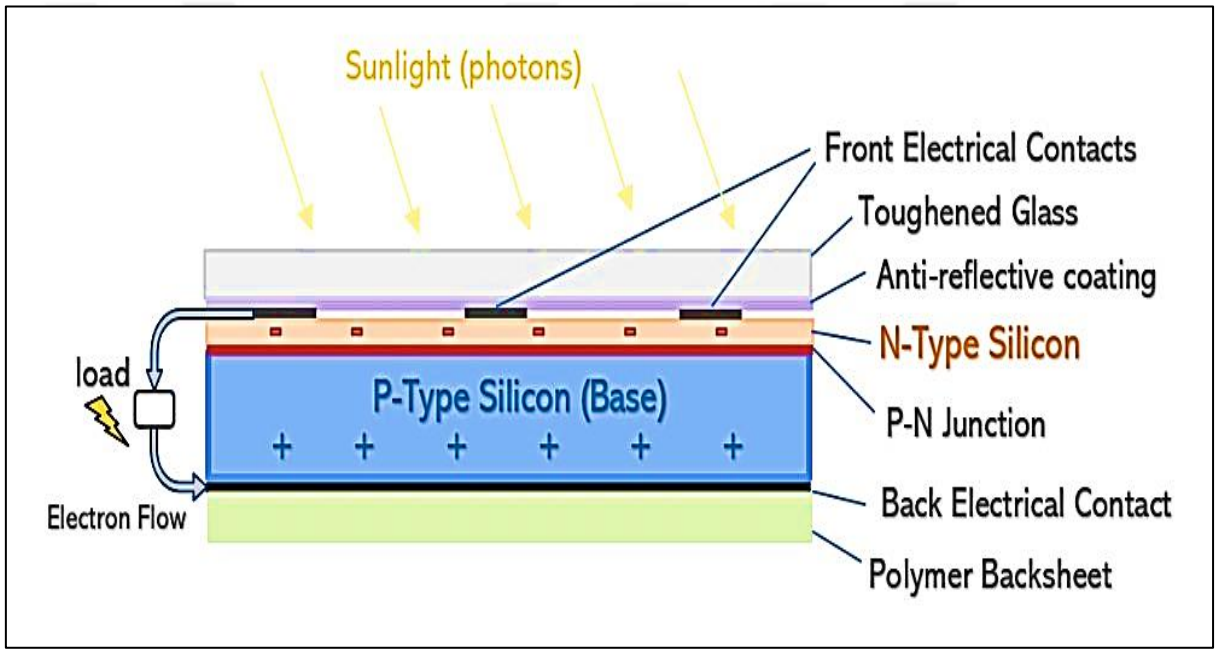

Figure 2.8: Structure of a photovoltaic cell [21]

When the two doped layers are introduced into contact, the extra electrons in the doped N cloth diffuse into the doped P material. The in the beginning doped N vicinity will become positively charged, and the at first doped P region is negatively charged. An electric field is thus created between them, which tends to push the electrons back into the N zone and the holes towards the P zone; a junction known as PN has been formed. By adding metal contacts to the N and P zones, a diode is obtained. When this diode is illuminated, photons having an energy ($h\,v$) greater than or equal to the bandwidth of the forbidden band EgExcite the silicon atoms and create positive and negative charges; Subsequently, electrons and openings made in

the P and N areas independently diffuse and show up at the space charge zone, where they cross the change zone, impelled by the inside electric field. The N region gets electrons and charges in a troublesome way. The P district takes openings and emphatically charges. If a charge is placed at the cell terminals, the electrons in the N-zone join the holes in the P-zone via the external connection, creating an electrical current [21].

The function of a cell can be represented by the curve I=f (V), which indicates the evolution of the current generated by the photovoltaic cell according to the voltage at these terminals from the short circuit to the open circuit.

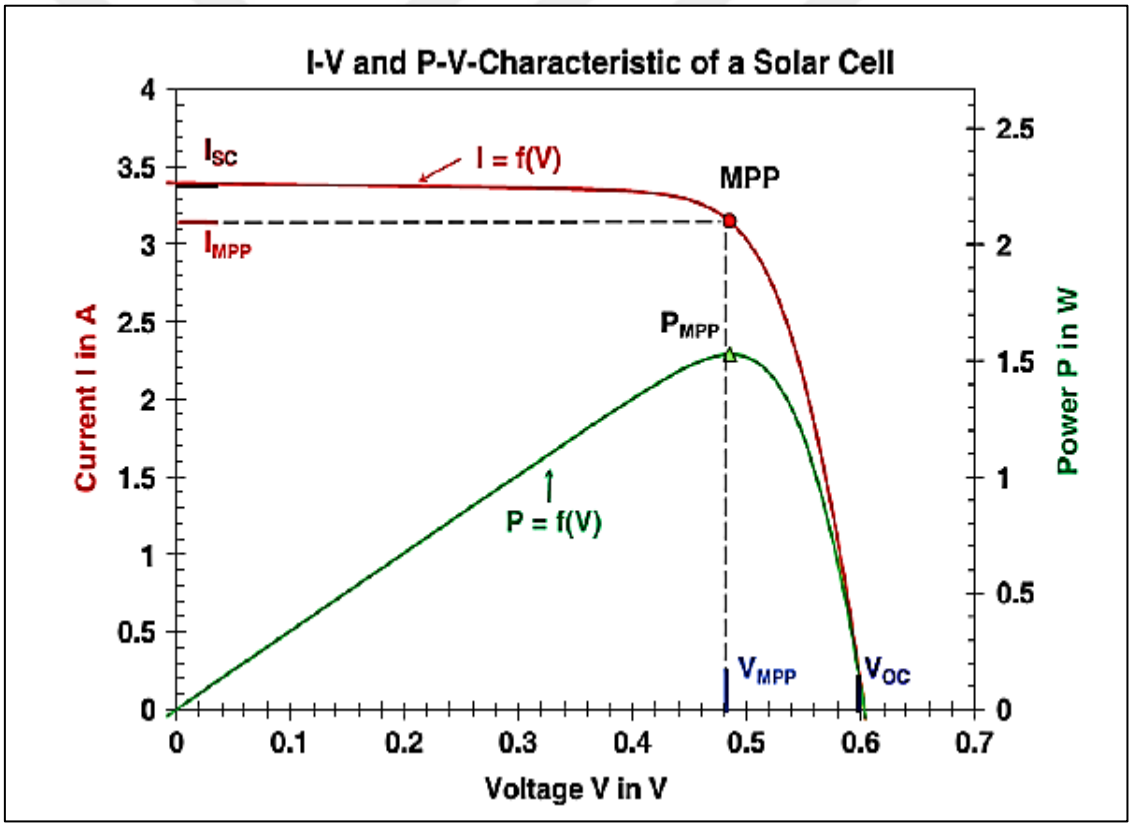

Figure 2.9: Current-voltage characteristic of a photovoltaic solar cell [18]

According to the current-voltage characteristic, it is possible to deduce the electrical parameters of the cell:

i- the short circuit current (Icc) corresponding to the current supplied by the cell when the voltage at its terminals is zero.

ii- the circuit voltage (Voc) corresponding to the voltage that appears at the terminals of the cell when the current flow is zero.

Between these two values, there is an optimum, at a voltage known as the maximum voltage Vm and a maximum current I'm, giving the greatest power ($Pmpp$) or peak power [21].

The pattern of the current-voltage characteristic figure 2.10 varies according to the environmental conditions (illumination and temperature)

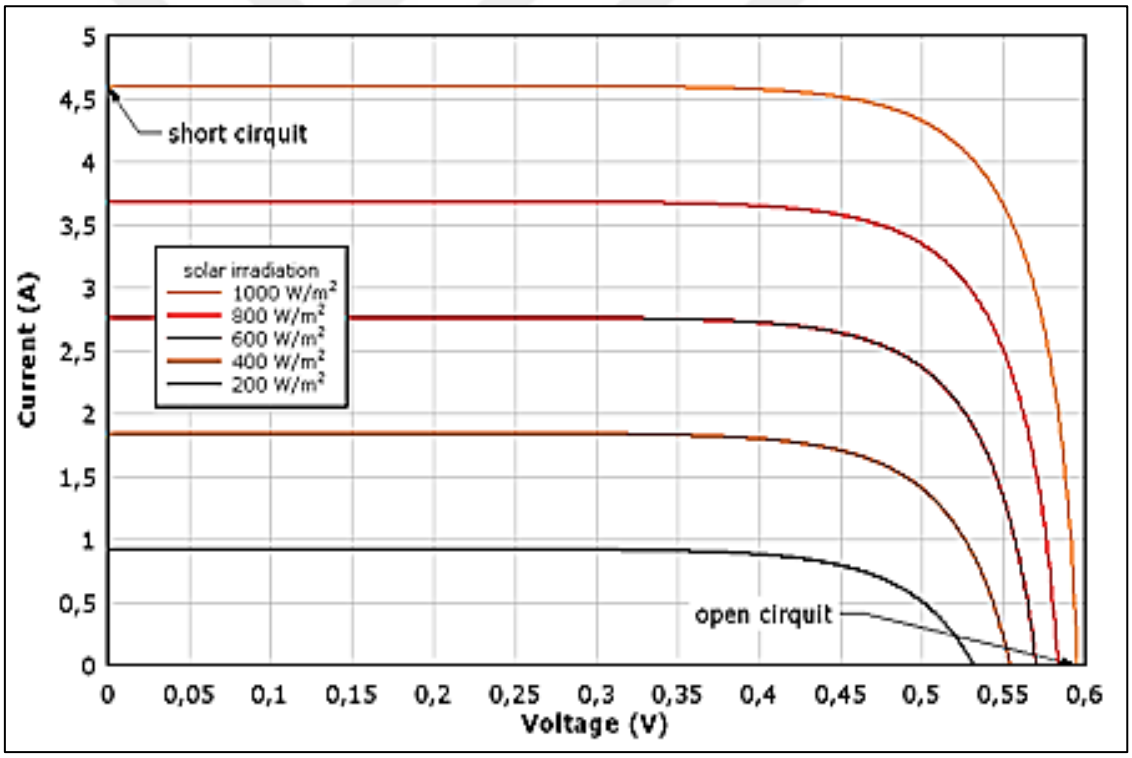

Figure 2.10: Influence of light on current-voltage characteristics of a photovoltaic cell [21]

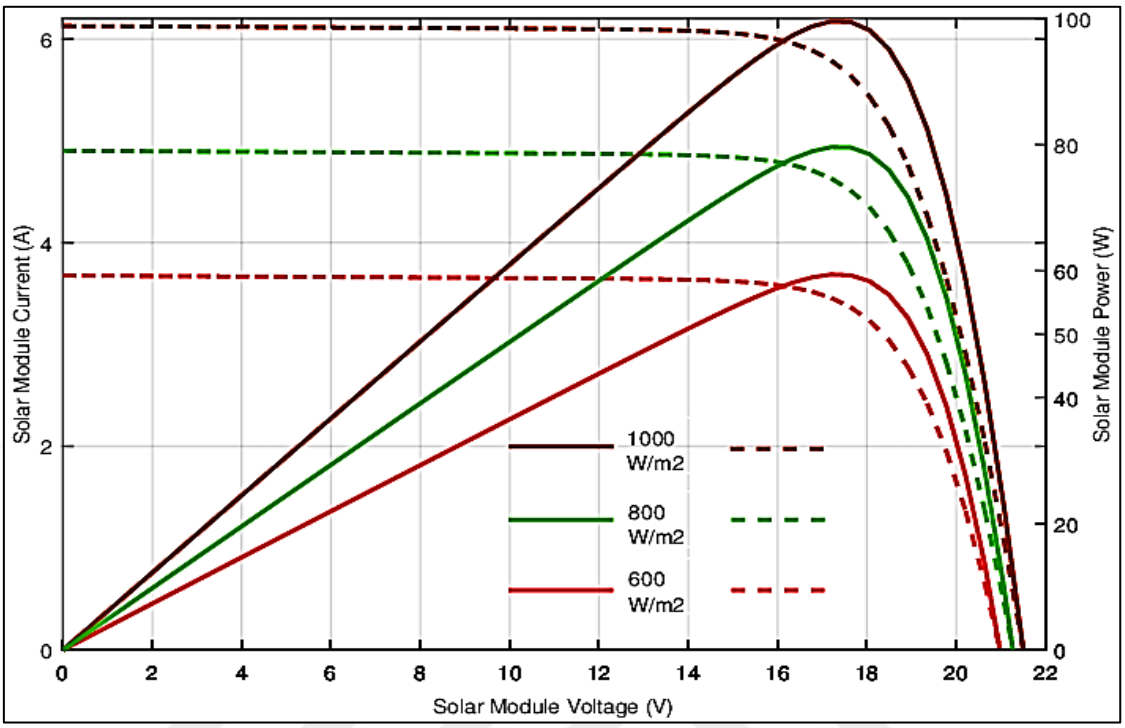

Figure 2.11: Influence of temperature on current-voltage characteristics of a photovoltaic cell [22].

The influence of temperature is significant on the current/voltage characteristic. The open circuit voltage decreases with the temperature increase, but the current varies very little with temperature. Therefore, the maximum power (Figure 2.11) delivered by the photovoltaic cell decreases.

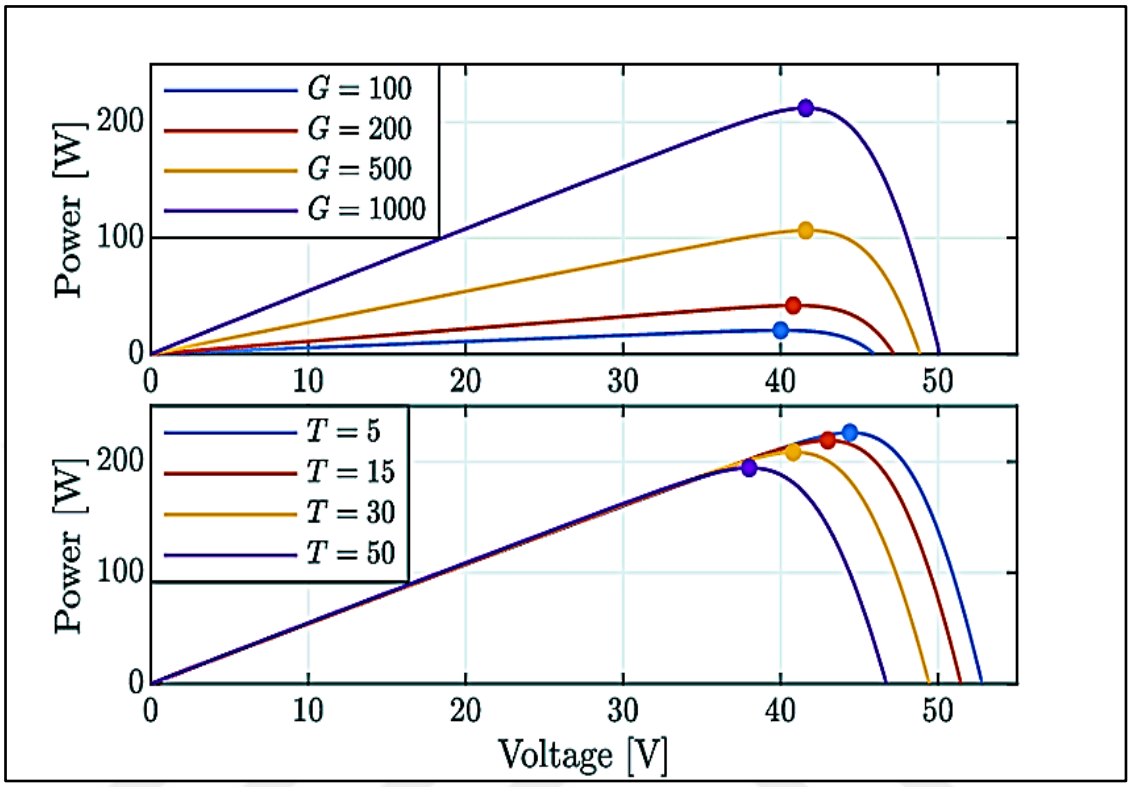

Figure 2.12: Influence of temperature on the power-voltage characteristic of a cell photovoltaic [22].

Unlike the variation in temperature, the variation in illumination affects the short circuit current that decreases as the illuminance decreases. The open-circuit voltage is not sensitive to this variation. This implies that:

i- The top-quality energy of the telephone is proportional to the illumination.

ii- The most electricity factors are at about the identical voltage (Figure (2.12)).

In this section, we have defined the majority of meteorological parameters, which are included in the calculations of solar radiation, whether out of the atmosphere or soil. We have recalled the most important empirical models of solar irradiation. Finally, we presented the experimental data, which are necessary for forming a database for our prediction models throughout this study. There are several ways to process solar radiation data. The next chapter is devoted to the state of the art of solar radiation prediction as that a time series

2.6 WHY PREDICT SOLAR RADIATION

The installation of any solar energy system in a given site requires preliminary studies. Indeed, sizing and simulation are essential to ensure optimal operation. To carry out such tasks, reliable measurements over relatively long periods of certain meteorological variables, and especially those of solar radiation, are essential. The lack of a long series of data or data series of low quality (discontinuity and unreliability) can combine errors in the design, sizing, and prediction of solar system performance, which hurts investment. Unfortunately, measurements of solar radiation are generally inaccurate and rare worldwide; Especially in Algeria, due to the high price of measuring devices. There are only a small number of solar radiation stations, which is why there is a lack of solar radiation measurements in large areas on the one hand. On the other hand, where these data exist, there are generally periods of failure due to failures or low monitoring, since the majority of these stations belong to establishments that do not benefit economically from these data.

Nonetheless, other meteorological boundaries, for example, temperature encompassing, dampness, wind speed are generally handily estimated in a bigger number of weather conditions stations with a moderately minimal expense contrasted with that of radiation. Then again, the enhancement of a planetary group or the reproduction of its exhibitions expect at any rate, day to day information, even, timetable of sunlight-based illumination. Consequently, it is alluring to foster connections between accessible meteorological information and sun-based light; and to foster exact strategies (models) to anticipate sun powered illumination. These models are tools that allow the generation of long series of solar irradiation data at different time steps; this could provide useful information to decision-makers in terms of selection, design, and planning of new solar power plants [19].

2.7 STATE-OF-ART ON SOLAR IRRADIANCE FORECASTING

This part gives an outline of the most normally involved systems in Sun-based Irradiance Guaging, as well as the ongoing status of the specialism of the models researched and thought about in this survey.

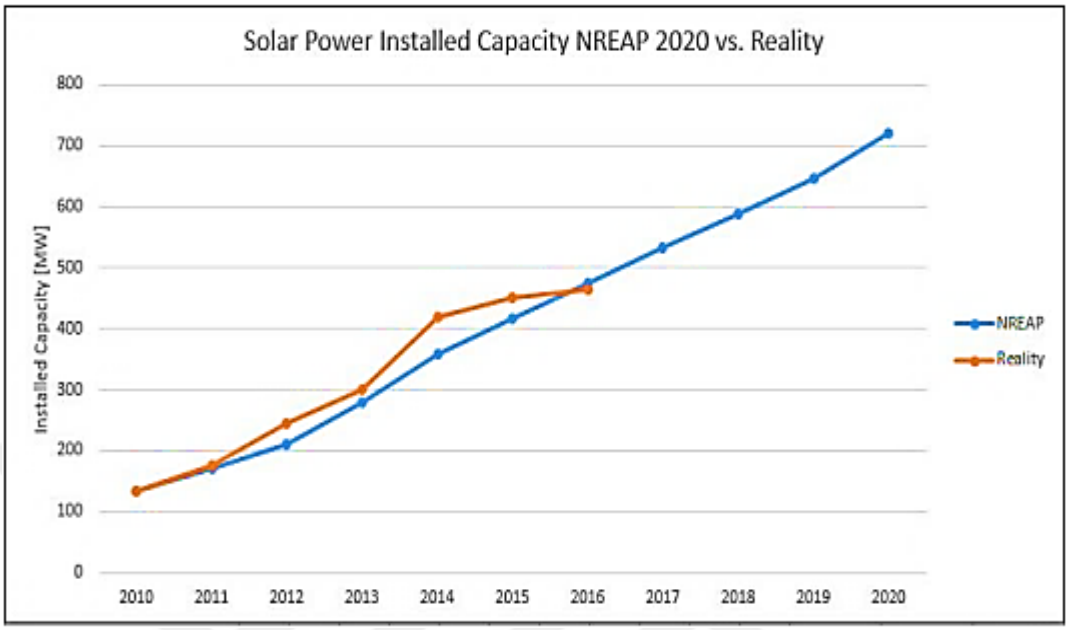

Figure 2.13: Comparison between NREAP and Reality Solar Power Installed Capacity from 2010 to 2020 [10][11].

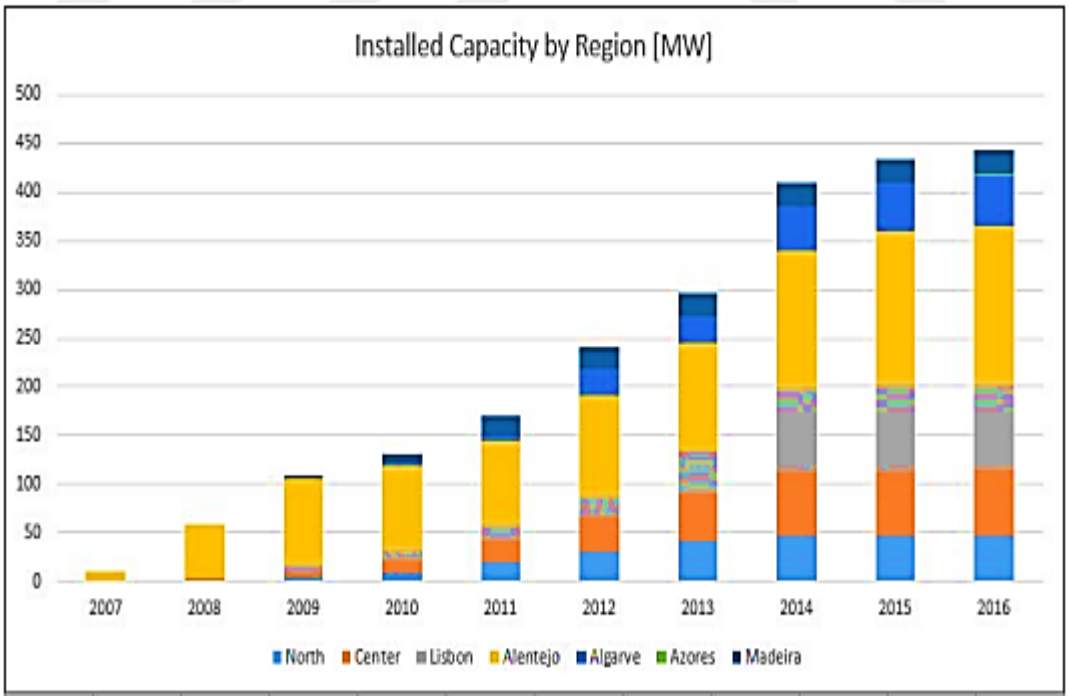

Figure 2.14: Solar Power Installed Capacity in Portugal by Region from 2007 to 2016. [11].

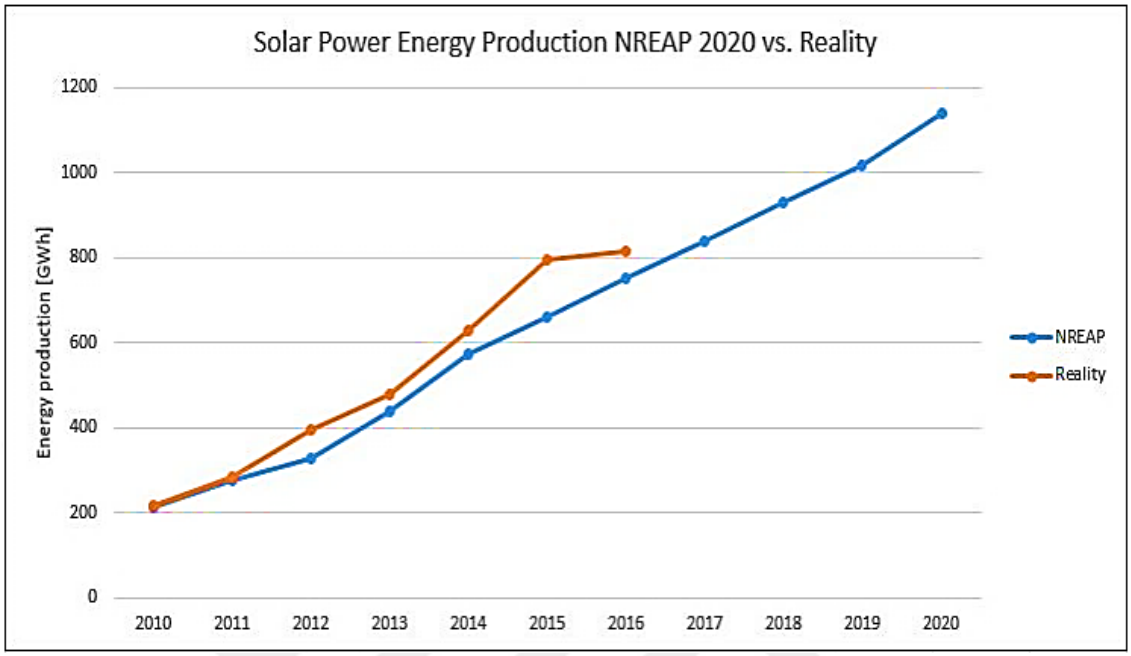

Figure 2.15: Comparison between NREAP and Reality Solar Power Energy Production from 2010 to 2020 [10][11].

.

A few techniques for determining sunlight-based irradiance and sun-oriented power have been created throughout the long term. The creators of [12] introduced a framework that joined a Help Vector Machine (SVM) Framework and a procedure of order of information series by meteorological sorts. The initial step is to order the PV power yield information into four gatherings (Overcast, Hazy, Stormy, and Radiant), and afterward the four SVM models will be made in view of the past four gatherings. It has likewise been endeavored to utilize a more stochastic model, as displayed in [13]. Certain Relative Mining (HISIMI) is a methodology that looks through the model's bona fide informational index for and utilizes the most similar information base cases to the ongoing case. All through this examination, different weight values are doled out in view of the degree of correlation between the legitimate and current cases. Following this examination, the model might handle the potential gains of the hopes of figure weakness and subsequently the point gauge for each future moment with the assistance of probability values and probabilistic abilities. In [14] the creators proposed one more blend of two unique strategies, this time the utilization of Wavelet Change alongside a Versatile

Neuro-Fluffy Surmising Framework. The wavelet creation is utilized to resolve the issue brought about by data discontinuities and non-periodicity in the contrast between them. The suggested assessment strategy begins by undermining the hidden information into unmistakable wavelet coefficient signals. Then, at that point, in the ANFIS procedure, these coefficients are utilized as information vectors, which helps new coefficients. To gauge future sunshine-based irradiance, these coefficients are recombined through a comparable wavelet change.

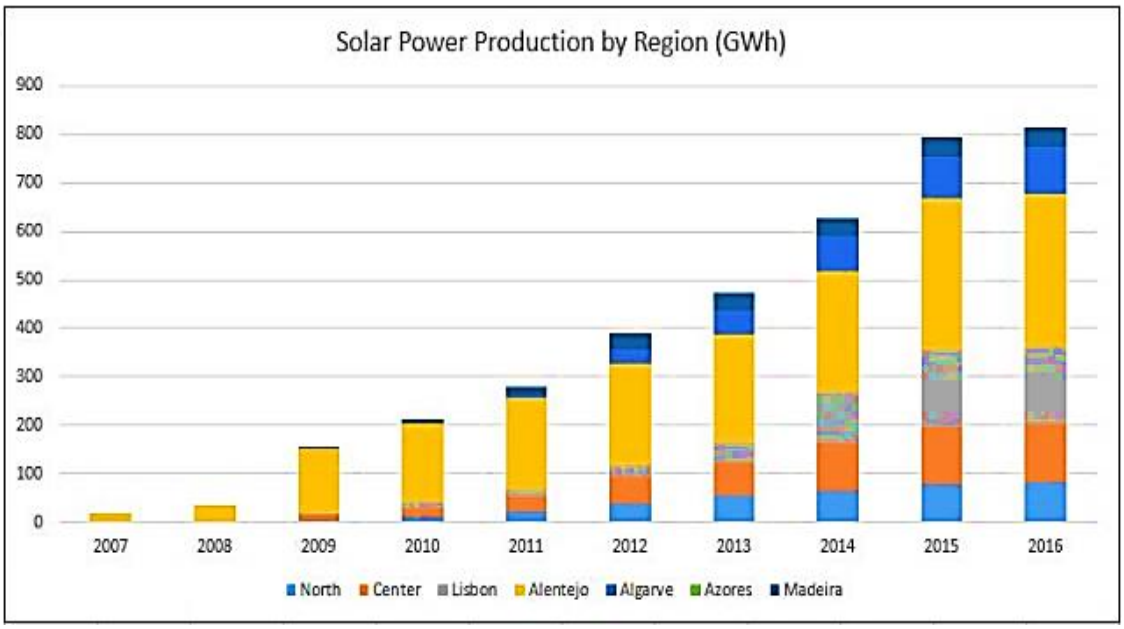

Figure 2.16: Solar Power Production in Portugal by Region from 2007 to 2016 [11].

The principal works in the subject of Brain Organizations happened in the late nineteenth and mid-20th 100 years because of an interdisciplinary joint effort between physicists, clinicians, and neurophysiologists. These distributions just incorporate general hypotheses of learning, discernment, or chiseling, with no particular mathematical models of neuron processes. These speculations put on weight when Warren McCulloch and Walter Pitts exhibited a state-of-the-art point of view on Mind Associations during the 1940s [15]. Their review shows the way that gatherings of fake neurons can enlist any number shuffling or significant capacity, which some see as the start of the mind network field. Notwithstanding, the chief utilization of a fake

mind network grew exclusively because of Obtuse Rosenblatt's production of the Perceptron Association and Related Learning Rule in the last part of the 1950s [16].

From that point forward, much work has been finished, and this innovation has started to be utilized in sun powered estimating. The creators looked at the presentation of three particular ANN Designs in [17]. The objective of their work was to decide the ideal design for making a 24-hour conjecture utilizing information from the earlier days' records from the day to anticipate, as well as information from that very days' and the following ones' records from the earlier year. They contrasted the FFNN engineering with networks with other exchange capabilities (Outspread Premise Capability Brain Organization, RBNN) or plans (Repetitive Brain Organization, RNN). They confirmed that the presentation of the two structures is practically identical, notwithstanding the way that the RNN and RBNN outflanked the FFNN for a couple of months.

The creators of [18] proposed and assessed the presentation of an ANN engineering (Multi-facet Perceptron) in guaging day to day light. This examination was led. They recently tried the strategies on non-fixed time series and found that the ANN and AR computations created more exact outcomes. They then, at that point, examined if a preprocessed data treatment would consider a huger supposition between the typical and assessed information. Utilizing just ANN, AR, and ARMA, the outcomes exhibited that this pre-dealt with treatment diminished check mistake by 5%-6% when contrasted with regular methods. The creators fostered an answer in that consolidates the notable ANN approach with a Wavelet Change in [19]. (WT). By separating the time-series information, WT assists with decreasing the erratic changes in sun-oriented irradiance. This proposed strategy emphatically upgraded the precision, productivity, and execution of existing ANN Structures. As should be visible from the information detailed in this paper, WT+RBFNN gave a lower figure to all seasons when contrasted with any remaining analyzed other options. Evelyn Fix and J.L.Hodges, Jr hypothetically proposed the Closest Neighbors Strategy in [20]. With the advancement of computer technology, it took around 15 years for Fix and Hodges' concepts to become a reality. The first generalized KNN decision rule was only investigated in 1970 by [21].

The creators displayed in [22] that, regardless of its ubiquity in grouping issues, the KNN strategy can likewise be utilized to manage relapse hardships. In this occurrence, they utilized the KNN to settle a precipitation and spillover examination of a given watershed. They thought about the adequacy of three particular occupation related strategies (the Sacramento Technique, ARMAX and the KNN). Breaking down the amount of lingering square blunders for the three strategies, they confirm that the Sacramento Strategy was plainly the most awful of the three models examined, and the KNN execution for this contextual investigation was truly similar to the ARMAX execution. [23] proposes a KNN variation that utilizes a Hereditary Calculation (GA) to assess model loads. The outcomes are contrasted with the ANN. It was scrutinized in a 24-hour energy cost foreseeing situation. The creators confirmed that consolidating the KNN with the GA is better than ANN. The KNN+GA approach has lower anticipating blunders. They utilized the KNN calculation to figure wind power in [24]. Likewise, to [23], the creators mixed two separate ways to deal with get a conjecture. In this situation, they utilized the KNN technique to find the k nearest authentic cases with qualities like the anticipated climate. They utilize a Piece Thickness Assessment approach in the wake of finding the k nearest cases to create a breeze power anticipated thickness. They showed that the proposed technique functions admirably for wind power anticipating, conveying either point or probabilistic gauges.

[25] made one more one more conclusive structure for the KNN approach. They determined GHI and DHI utilizing close by irradiance information and sky pictures in this review. They photo the sky while driving the camera toward the pinnacle, filter out unessential data (e.g., trees, plans), and convert the 8-cycle RGB channels of the handled pictures into drifting focuses for use in the expected structure. The outcomes show that the KNN model accomplishes huge figure upgrades over a reference determination estimate, at the same time, in opposition to assumptions, the presentation of sky imagens into the KNN model doesn't increment execution fundamentally.

2.8 RELATED WORK

The main efforts to anticipate irradiance were made a long time back (Detlev Heinemann) Following years saw just minor endeavors or progress in the improvement of sun powered irradiance guaging strategies. Sun oriented power guaging is subject to variables like the condition of the environment, information on the sun's way, and the qualities of a sun-based energy plant, which utilizes the sun's energy to create sun powered power. The ARIMA model, Counterfeit Brain Organization (ANN), and Mathematical Climate Expectation are the most normally involved strategies for anticipating irradiance (NWP). Measurable techniques perform well for exceptionally transient determining in the time scope of 30 minutes to 6 hours. NWP is used for forecasting up to two days in advance. As the forecasting window lengthens, the accuracy of statistical techniques degrades rapidly. Absolute sky imager approaches are used for transient guaging, considering cloud data. The Cloud Movement Vector (CMV) is assembled and traveled through time, and sun-based light from the cloud is recovered thusly utilizing mathematical reasons and semi-observational models. The period of time for these not entirely settled by the speed of the cloud. The semi-observational heliostat strategy was utilized to remove data about cloud structure from satellite pictures (Annette Sledge, 2003). The CMVs are then developed by looking at two back-to-back pictures, as in sky imager-based methods. Utilizing CMVs, the cloud data is pushed ahead in time prior to going through a last smoothing stage. The heliostat technique is utilized to recuperate sun-based light. NWP is utilized to estimate for as long as six days or more. These procedures are hard to apply and are now and again unequipped for evaluating vulnerability in expectations. It is hard to add a vulnerability model to the heliostat approach consistently. NWP models can deal with vulnerability, however at a high computational expense that required the utilization of committed supercomputers. Among the strategies depicted above, the ARIMA model can produce mistake bars for expectations, which helps with evaluating possible dangers. The satellite-based model is more dependable. The clearness file is utilized in this model to represent the Sun's development impact on the satellite noticed sun-based irradiance field, which is taken out by separating it with an unmistakable sky model. The non-straight, non-parametric method Gaussian cycle relapse can be utilized to get familiar with the elements of

this model (GPR). The recursive Gaussian interaction is Bayesian, giving best gauges as well as probabilistic conjectures.

2.8.1 Forecasting techniques

Forecasting solar irradiation components is the practice of anticipating various sun irradiation components such as global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI) for a certain PV site in advance. Forecasting such components in advance for a specific place, however, is a difficult endeavor due to the effect of fluctuating meteorological circumstances. This section discusses the key strategies for developing forecasting models for predicting solar irradiation components. In the literature, various methods for solar forecasting have been addressed, which can be broadly classified based on data sets, model structure, and operation. Figure 2.17 depicts the various types of data set-based and structure-based approaches.

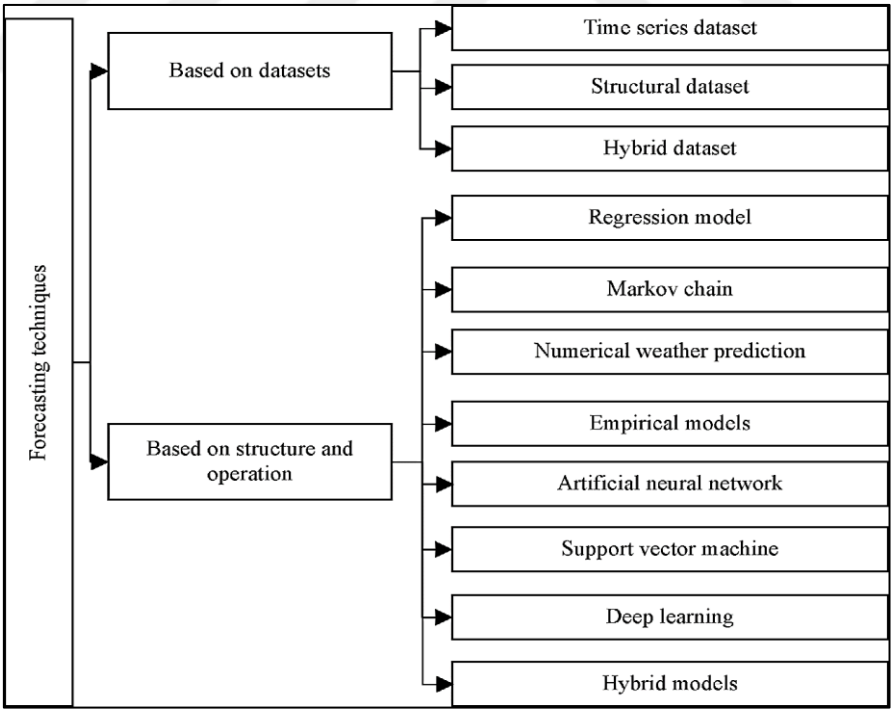

Figure 2.17: Types of forecasting techniques based on data processed and their structure

There are a few models open in the writing that vary in design, working, and usage. These models are extensively ordered into three kinds: factual [34], physical [35], and crossover [36,37]. These models were made utilizing different procedures, for example, savvy determining strategies, relapse, support vector machines (SVM), a high-level relapse philosophy, Markov chain, wavelet change based insight methods, numerical procedures, and mathematical climate expectation (NWP).

Relapse models depend on the methodology of fostering a numerical connection among reliant and free factors [38]. Notwithstanding, different models in light of essential direct relapse and various straight relapse approaches have been created by scientists, however autoregressive models are definitely more pervasive than basic and numerous straight relapse. The connection between's the reaction and indicator factors is estimated utilizing autoregressive models [6]. Furthermore, various categories of time series have already been formed based on linearity/nonlinearity and stationarity/non-stationarity, such as autoregressive (AR), moving average (MA), autoregressive and moving average (ARMA), autoregressive integrated and moving average (ARIMA), seasonal autoregressive integrated moving average (SARMIA), autoregressive integrated and moving average with explanatory variables (ARIMAX) [27]. In 1992, [28] offered two new methodologies for analyzing stationarity, the first termed MA and the second AR [40]. ARIMA is the most often used model for determining the relationship between predicted and measured outcomes.

McCulloch and Pits [29] introduced the first artificial neural network (ANN) concept in 1943. The neuron in the human brain performs many types of parallel processing and pattern recognition analyses [30]. The same phenomenon can be used to tackle nonlinear mathematical problems like pattern recognition, forecasting, and image processing, among others. This technique-based model must be trained numerous times to achieve the appropriate value to weight for mapping the input and output [31]. To expect the result factors, the ANN utilizes numerous strategies like the Levenberg-Marquardt (LM) calculation, the scaled form angle, and the Pola-Ribiere form inclination [32]. Moreover, [33] utilized different blends of information boundaries and qualities-based bunches of info information for expectation.

[40] created the LSTM network to forecast solar irradiance for the next hour. They forecasted using the weather data collected from the island of Cape Verde, Santiago. Based on different data set sizes, they compared the created model to the BPNN and the persistence approach. Using LSTM and 2 years of data, the model accuracy was improved by 18.34%. However, with 9 years of training data and 1 year of validation, it exhibited a 42.9% improvement in RMSE. [41]. Despite the fact that [106] utilized LSTM to gauge 2, 6, and 24 hours ahead for four spots in the Thar Desert, the MAPE acquired went from 6.79% to 10.47%, while the RMSE went from 0.099 to 0.181 for various designated regions [42]. Nonetheless, temporarily skyline, [43] proposed a gated repetitive unit network-based PV yield power estimating. The Pearson coefficient was used in the model to remove the info factors that influence PV estimates. By gathering the sets in light of similar qualities, the k-implies bunch method was likewise used to diminish preparing time. Each group was then associated with the GRU network to estimate the PV yield power. The suggested model's performance was tested using RMSE in contrast to existing models such as LSTM, BO, SVM, and ARIMA. The GRU-based network achieved an average RMSE of 0.036, which was better than the above-mentioned models [44], which used LSTM and the k-means algorithm for short-term solar irradiation prediction under complex meteorological conditions. On their clearness index, they split the days into three categories: sunny, cloudy, and mixed. The LSTM algorithm was applied to the input data based on the clear sky index, and then all of the values were ensemble to generate the final GHI forecast. With RMSE going from 41.37 to 66.69 (W/m2), MAE going from 30.19 to 46.04 (W/m2), and R2 going from 0.95 to 0.97 [45], the proposed model beat CNN and SVM. [46] then again estimated present moment sun-based PV power utilizing various GRUs. To work on the nRMSE, various GRU networks were connected in a fountain structure. The primary organization's result turns into the contribution to the following organization, etc until the nRMSE approaches a base worth. As far as nRMSE, the model precision was then contrasted with that of the SVR, KNN, and regular GRU. With three organization associations, the model outperforms others with a nRMSE of 9.64% [47].

All in all, unique mixture models have been proposed by different researchers. Notwithstanding, picking a proper model from among all models for a particular issue is very

troublesome on the grounds that each model was created for an alternate locale, for an alternate arrangement of information factors, for various natural circumstances, for various preparation information, and for various mistake frameworks. In any case, for information pre-handling, wavelet change (WT) beats observational model disintegration (EMD), though GA/PSO can advance the results better compared to other people. The consideration system, related to LSTM, can be used to predict a super brief time frame skyline, which performs all the more precisely. The significant spatial and worldly changeability in the information can be tended to by CNN and GRU hybridizations. Thus, as a rule, the crossover design of profound learning models outflanked others.

Table 2.1: Comparison table

Ref.	Region/place	Lat/Long	Timeahead	Training period	Testing period	Input variable	Outpuvariable	Technique	Error parameters	Error/%
[31]	Turkey	36°/42°		4 years,train 9 station	3 station	Lat, Long, Al, M, MSD, T_m	SR	ANN with SCG, CGP, LM	MAPE, R^2	MAPE < 6.78%, R^2 = 99.7768%
[32]	Helwan, Egypt			1 year	1 year	RH, T, WS, WD,	UV, IR, GI	ANN	RMSE	RMSE: IR = 5.02%,
						CC				UR = 7.46%,
										GI = 3.97%
[33]	India: Ahmadabad,			10 stations	10 stations	Lat, Long, Al, t,	CI	ANN-FFBP	RMSE	RMSE = 4.5%
	Nagpur, Mumbai,			in India	in India	M, T, RH, R, WS, LW				
[34]	Dezful, Iran	32°.16 N 48°.25′ E	24 h	1398 d	214 d	DAT_m, RH, SD, EWS	DGSR	MLP, RBF	MAPE	MAPE = 5.21%
[36]	Centra Queensland	23°.38′/150°.58′	3 h	4 m	1 m	AT, WS, WD,	WE, SE	ANN	R	R = 0.96399

2.8.2 Summary

Sun oriented light is intensely impacted by topographical and climatic variables. Sunlight based irradiance's dynamic way of behaving straightforwardly affects the unwavering quality of PV incorporated frameworks, the energy market, and power utility organizations. Thus, to run the power framework easily, an exceptionally careful and reliable sun-oriented irradiance expectation is required. Be that as it may, exact determining of sunlight-based irradiance is one of the most troublesome issues to finish, with various frantic models previously settled in the writing.

3. PROPOSED SYSTEM

With the coming of minimal expense computerized photography, the strategies used to detect the world's climate have gone through a change in outlook. Pictures got from geostationary satellites were generally the essential wellspring of visual data about the world's climate. These satellite pictures oftentimes have low worldly and additionally spatial goals, making them less valuable in nations like Singapore, where climate peculiarities are regularly exceptionally confined. High-goal WSIs can give more definite neighborhood data about cloud arrangements in these situations. With a hemispheric focal point, these gadgets catch the whole sky from the beginning customary stretches. The subsequent pictures have a higher goal than satellite pictures. Besides, the camera's vertical direction permits it to catch low-lying mists.

GHI is defined as the sum of diffuse radiation incident (DRI) on a horizontal surface plus direct normal irradiance (DNI) projected onto the horizontal surface, which can be expressed as follows [20]:

$$GHI = DRI + DNI * cos(z) \tag{3.1}$$

where z is the solar zenith angle of the Sun. We have found tremendous articles that tackle the solar irradiation forecast using Artificial intelligence models. Before, going through these models, it worth highlighting that solar irradiation models can be divided in 3 main categories [7][15] which are:

i- Numerical Weather Prediction models which are deterministic models, based on current weather conditions [25]. This kind of model gives better prediction from 6 hours on-wards [7].

ii- Statistical models, which are based on the on-site time-series measurements. This kind of models have better performance for prediction up to 6 hours [7].

iii- Hybrid models, is yet another kind of models that combine different algorithms from the previously cited models, in order to improve the prediction accuracy. We can cite these references [5], [4] for more details.

Most of the papers make a comparison with a simplistic model that is Persistent model. its also known as naïve model [8]. It is recommended to use it as a comparison before implementing complex models. The persistence model supposes that global irradiance at time t is best predicted by its value at time $t-1$:

The Clear Sky model is a deterministic model that estimates global irradiance in clear sky conditions at any given time. A cloudiness index or clearness index is also calculated using clear sky models [14]. There are many different types of clear sky models, ranging from simple models that are only a function of the solar zenith angle to more complex models that incorporate many atmospheric parameters, such as aerosols and precipitable water [14].

In the recent decade, the Artificial Neural Networks (ANNs) models used successfully to represent complex functions. They have been used successfully as well for GHI predictions. These models can utilize time series information, time series pictures, or both. Mellit and Pavan [16] proposed a Multi-facet Perceptron (MLP) that can foresee GHI for the following 24 hours. It acknowledges mean everyday irradiance and mean day to day air temperature as information boundaries; comparably, [12] utilized a multistage ANN to anticipate GHI for the following day. The organization's feedback information is the typical barometrical strain anticipated by another ANN, as well as different climate information from the earlier day. [22] involved an ANN to foresee GHI hourly qualities in a solitary step.

3.1 THEORETICAL FRAMEWORK

The prediction of incoming solar energy is becoming more important as solar power generation increases. Since photovoltaic (PV) energy age is straightforwardly connected with sunlight-based radiation, it is a significant boundary in sun powered energy applications. Sunlight based radiation differs nonlinearly because of climatic occasions like shady climate, downpour, stickiness, etc. Thus, assessing sun-oriented radiation is a basic issue in the sun

powered energy field. By 2020, the Indian government expects to boost the utilization of sustainable power. At present, the nation doesn't have its own guaging model. Weather conditions satellite information will further develop estimating for any district. Many methods are available for forecasting short-term solar irradiance, but according to the literature, a satellite-based data-driven model is superior. We used the clearness index as an input to keep things simple. The clearness index can be used to calculate solar irradiance. Forecasted solar irradiance can be used to estimate power. The Gaussian process regression model was used in this study, and the PV estimation method was used for power forecasting.

3.2 PROPOSED SYSTEM

With the surge of perturbations that climate change has brought in the last years, the need for energy diversification has risen. Better solar irradiance forecasting would ensure efficient management of photovoltaic energy, i.e storing energy while in maximum solar irradiance.

The present paper describes a study that was made on satellite imagery (GOES-13)[7] measured by seven SURFRAD [2] stations in the continental US. The goal was to use satellite imagery in order to build a robust predictive model that would provide Global Horizontal Irradiance (GHI)[26] values at specific points on a given map for both present and future times. Specifically, the prediction horizon interest is $T0$, $T0 + 1h$, $T0 + 3h$,$T0 + 6h$, i.e, predicting GHI values for a specific timestamp as well as for 1, 3, and 6 hours in the future. Such models can have a significant impact on the future of large-scale renewable energy plants as it is critical to power system planning, management, and operations. Many existing GHI prediction solutions rely on physical and mathematical models. However, these solutions do not always capture complex temporal phenomena (clouds, winds, etc.). The hypothesis of our project is that satellite imagery for nowcasting can be used to help model these complex phenomena and seek to develop models leveraging open data sources. As for the metrics we used to evaluate our model, we used Mean Squared Error (MSE) as loss function to optimize our model and kept track of the Root Mean Squared Error (RMSE) [8] to evaluate the performance of our model.

Solar Irradiance : power per unit area (watt per square meter, W/m^2), received from the Sun in the form of electromagnetic radiation"

Using satellite imagery and other metadata such as datetime, station ID, etc., predict GHI values up to 6 hours in the future at specific points on the map.

$$T = T_0, \quad T = T_0 + 1h, \quad T = T_0 + 3h, \quad T = T_0 + 6h. \tag{3.2}$$

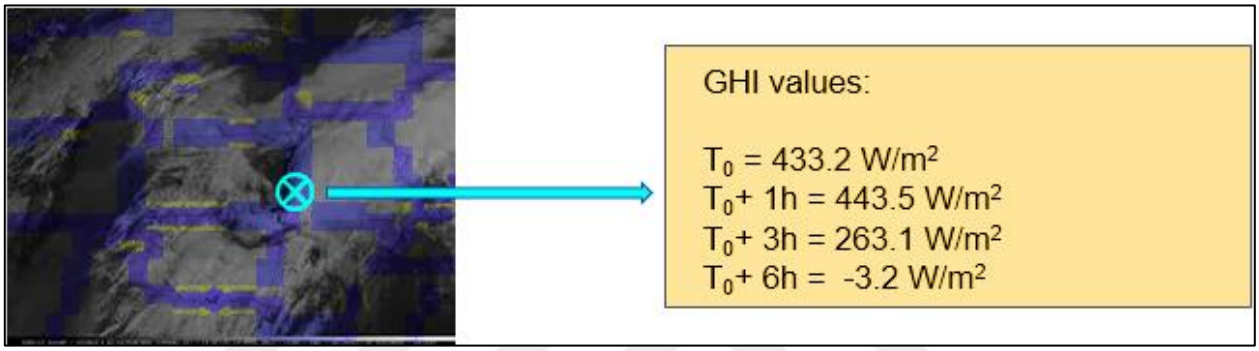

Figure 3.1: GHI value using satellite image

3.3 METHODOLOGY

3.3.1 Dataset

The NSRDB dataset contains both noticed climate information (temperature, pressure, overcast cover, sun peak point, etc) and sun-based power information (watts per square meter). A few suns powered radiation estimations are remembered for the assortment, including Diffused Typical Irradiance (DNI), Diffused Even Irradiance (DHI), and Worldwide Level Irradiance (GHI) (GHI). We utilized GHI information since they incorporate DHI, DNI, and encompassing sun powered energy reflected from encompassing surfaces. Therefore, it is a reasonable sign for sun powered charger readings. The NSRDB data is measured once every 10 minutes. We investigate a single location - Las Vegas, Nevada, USA. We limit our analysis to the year 2019 only. Our dataset contains more than 50,000 distinct observations, each with 16 features (including Time values such as Month, Day, Hour and Minute) shown in table 3.1

with corresponding measure of GHI. We obtain the sunrise and sunset times for each day from Sunrise-Sunset-API [2]. Using this, we add a new boolean column 'isDay'.

Table 3.1: Time values

Month	Day	Hour
Minute	Temperature	Cloud Type
Fill Flag	Surface Albedo	Ozone
Pressure	Dew Point	Precipitable Water
Wind Direction	Wind Speed	Relative Humidity
Solar Zenith Angle	isDay	

3.3.2 Data Processing

The first step of our data processing was to handle NaN values. In fact, we converted string nans to numpy NaN and for each station that had missing nightly clearsky GHI or true GHI, we filled those with 0 since there is no solar irradiance at night. Then, we interpolated missing GHIs based on both previous and future values in the dataframe. For the reasons discussed above in the data analysis section, we dropped the dataframe observations where there was no path to the NCDF files and for the .H5 files. Then, we shuffled the dataframe so our model could remain as general as possible.

Having preprocessed the data frame, the next step was to crop the images for training and validation set. Our training set spanned from 2010 to 2014 and our validation set consisted of the images from 2015. The goal was to crop the GOES-13 images centered on the coordinates of the stations so our data loader would be more efficient. Before doing so, we needed to map the station coordinates to their pixel value. In fact, while we have the station coordinates, e.g.,

BND which sits at 40.05192, -88.37309, this does not correspond to the pixel 40.05192 in our image which has a shape of (650, 1500). The cropping experiment and mapping the stations coordinates on the images are shown in our Image_analysis.ipynb notebook. After this mapping was applied, we observed that one particular station, FPK, is located at the 607th pixel on the x-axis. It is an important information that we considered further along our data processing when defining the cropping window hyperparameter, otherwise the crop would be out of bound.

3.3.3 Data Analysis

We were provided with two main data sources, such as GEOS13 imagery and some metadata stored as a Pandas data frame [19]. Since the raw NetNCDF imagery of the Geostationary Operational Environmental Satellite provided by the National Oceanic & Atmospheric Administration (NOAA) was over 800GBs, it was split into 15-minute chunks. More specifically, it was provided in three different versions.

The original, 16-bit, GZIp-compressed (lossless) NetNCDF files, in 15-minute chunks.

Repackaged, 16-bit, JPEG2000-compressed (lossy) HDF5 archives, in 1-day chunks.

Repackaged, 8-bit, JPEG-compressed (lossy) HDF5 archives, in 1-day chunks.

Each GOES-13 satellite imagery has five channels, each one of them corresponding to different wavelength. Keep in mind that the satellite is geostationary, i.e it is stationary in relation to a fixed point on Earth. As for the metadata, each column in the dataframe specifies an attribute for every 15-minute entry. Some of these attributes include the absolute path on the remote server (Helios) to the GEOS-13 imagery files, binary flag indicating whether it is day time or not at the station, a cloudiness categorical flag indicating the weather conditions at the station, the clearsky GHI [20], and the real measured GHI at the station.

By doing exploratory data analysis, after remapping string nans to numpy NaNs, we observed that there were some missing values for certain features (Table 3.1). Note that the total number of observations is 210 336.

Table 3.2: Relevant percentage of missing values per feature over the whole DataFrame

Features	Missing percentage
ncdf_path	17.4098%
hdf5_8bit_path	0.015%
hdf5_16bit_path	0.015%
BND_GHI	0.086%
TBL_GHI	0.2444%
DRA_GHI	0.5980%
FPK_GHI	0.2696%
GWN_GHI	1.2242%
PSU_GHI	0.1084%
SXF_GHI	0.362753%

Based on the utility module provided, we observed that even if the percentage of missing hdf5_path files was lower than the percentage of missing ncdf_path files, we should not take the occurrences where there are hdf5_path available but no ncdf_path available because on such occurrences, the hdf5 files may correspond to another timestamp where the data was available for both files. Then, we explored if the distribution of nighttime and daylight observations were equal for all of the seven stations. It turns out that effectively, the distribution of nighttime and daylight observations is almost equal for all stations (around 50-50). We also wanted to confirm our hypothesis that the location of the station would affect the daylight GHI values. Indeed, our hypothesis was valid, as a station like DRA, (Figure 3.1)

which is in Nevada receives more sunlight than a station like PSU (Figure 3.2) which is located in a northern state (Pennsylvania).

Next, we wanted to verify if the clearsky_ghi values provided in the data frame were hard limits or if the true GHI values could exceed them. As observed, GHI values are not bounded by clearsky_ghi values (Figure 3.3) This gave us a good indication to not bound our predicted values by the clearsky GHI. As per the underlying structure of the data, it is important to note that an image is 650 pixels per 1500 pixels for each of its channels, that one pixel is equal to 16km2 and that each GOES-13 image consists of five channels.

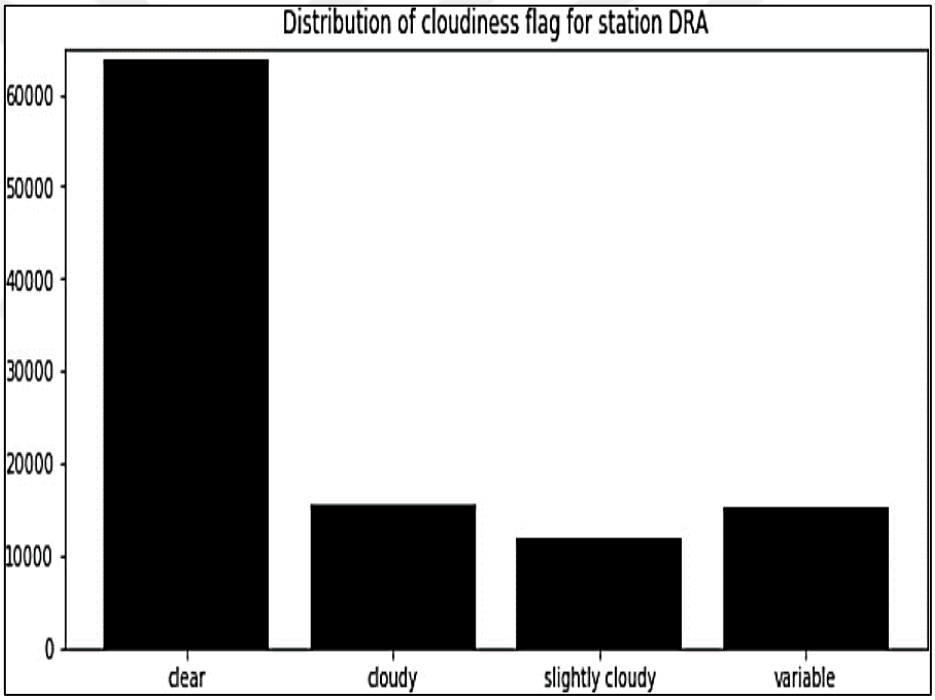

Figure 3.2: Cloudiness distribution for DRA station

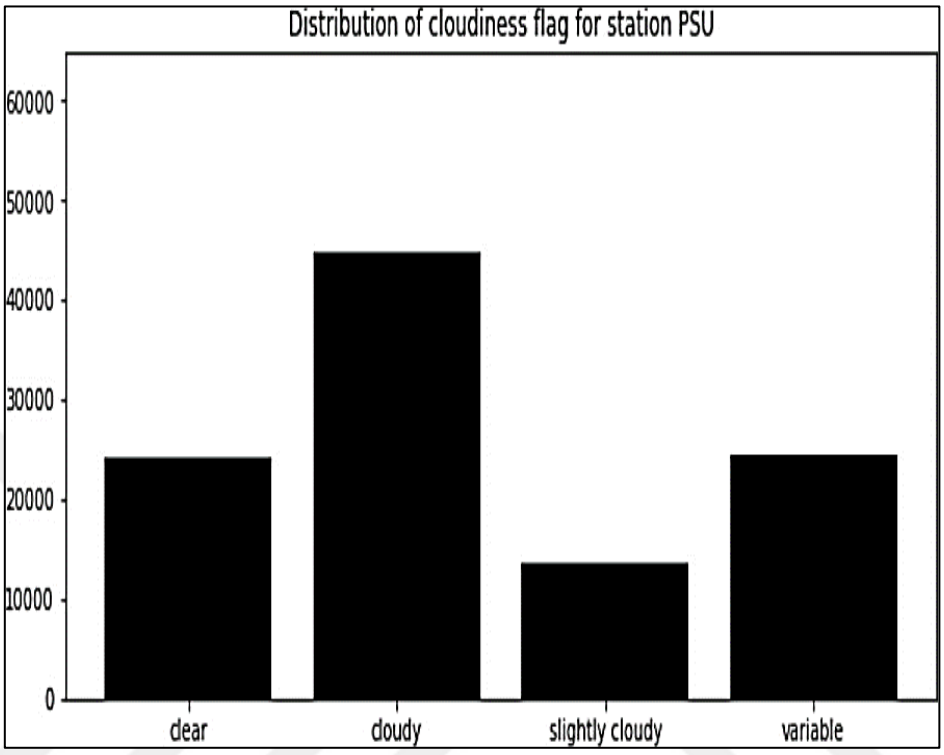

Figure 3.3: Cloudiness distribution for PSU station

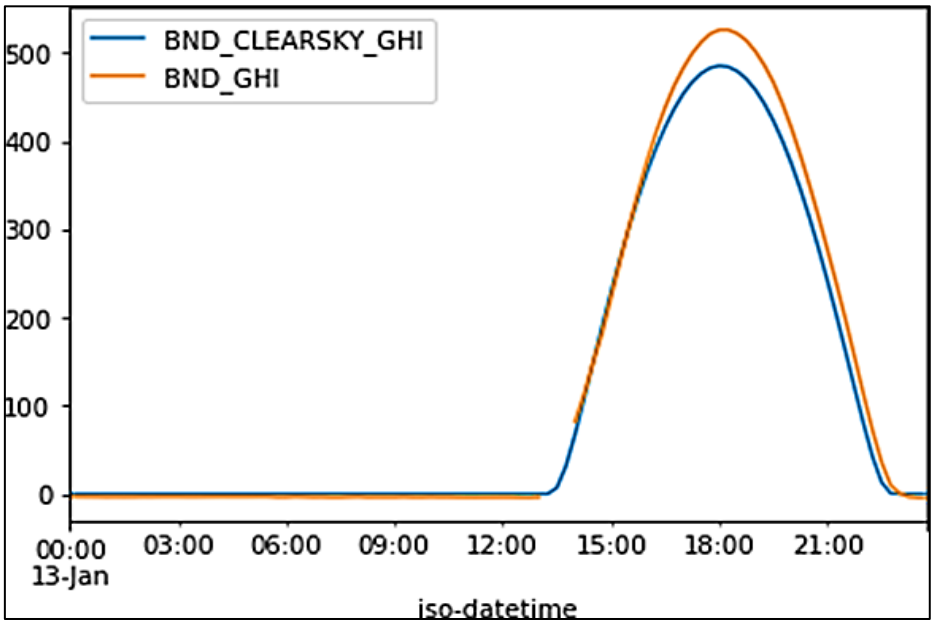

Figure 3.4: True vs clearsky GHI values at the BND station for one day

3.3.4 Handling Missing Data

As shown in Table 1 we have missing GHI values during daytime along with missing images. We have implemented methods to handle this issue [1]. To summarize these experiments, the GHI value was interpolated with two methods. First, we used the nearest available values from previous and future timestamps and interpolate linearly all missing values in between. This approximation is good enough for little gaps of a couple of timestamps of 15 minutes. But one it comes to big gaps of more than 3 hours, then the linear interpolation is not any more realistic. To handle this situation, we have used the GHI values of the day before and the day after to calculate the meaning of GHI value for each missing timestamp. This method gives a better approximation that linear interpolation as it makes a smoother transition from the day before to the day after.

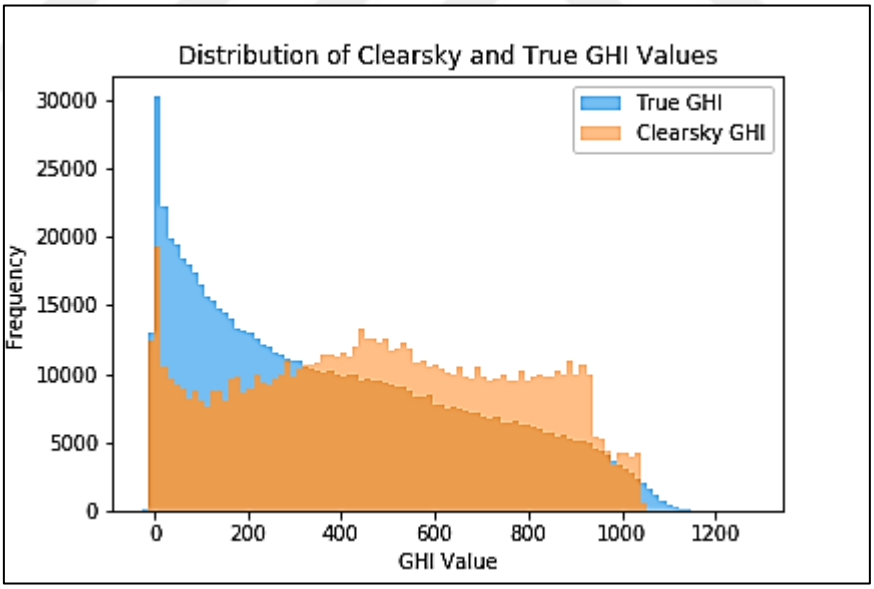

Figure 3.5: Daytime GHI Distributions

Regarding the images, we have used the same principle, that is taking the linear interpolation for the missing image from the previous timestamp and the next timestamp. As shown in the

notebooks, the interpolated images look realistic. We discuss in more detail the methodology adopted for handling missing data in the Data processing section.

3.3.5 Dealing with NaNs

We interpolated missing GHIs using a linear interpolation based on the previous and next occurrence of valid data. We also dropped night-time observations for training, replaced string nans, shuffled the data frame.

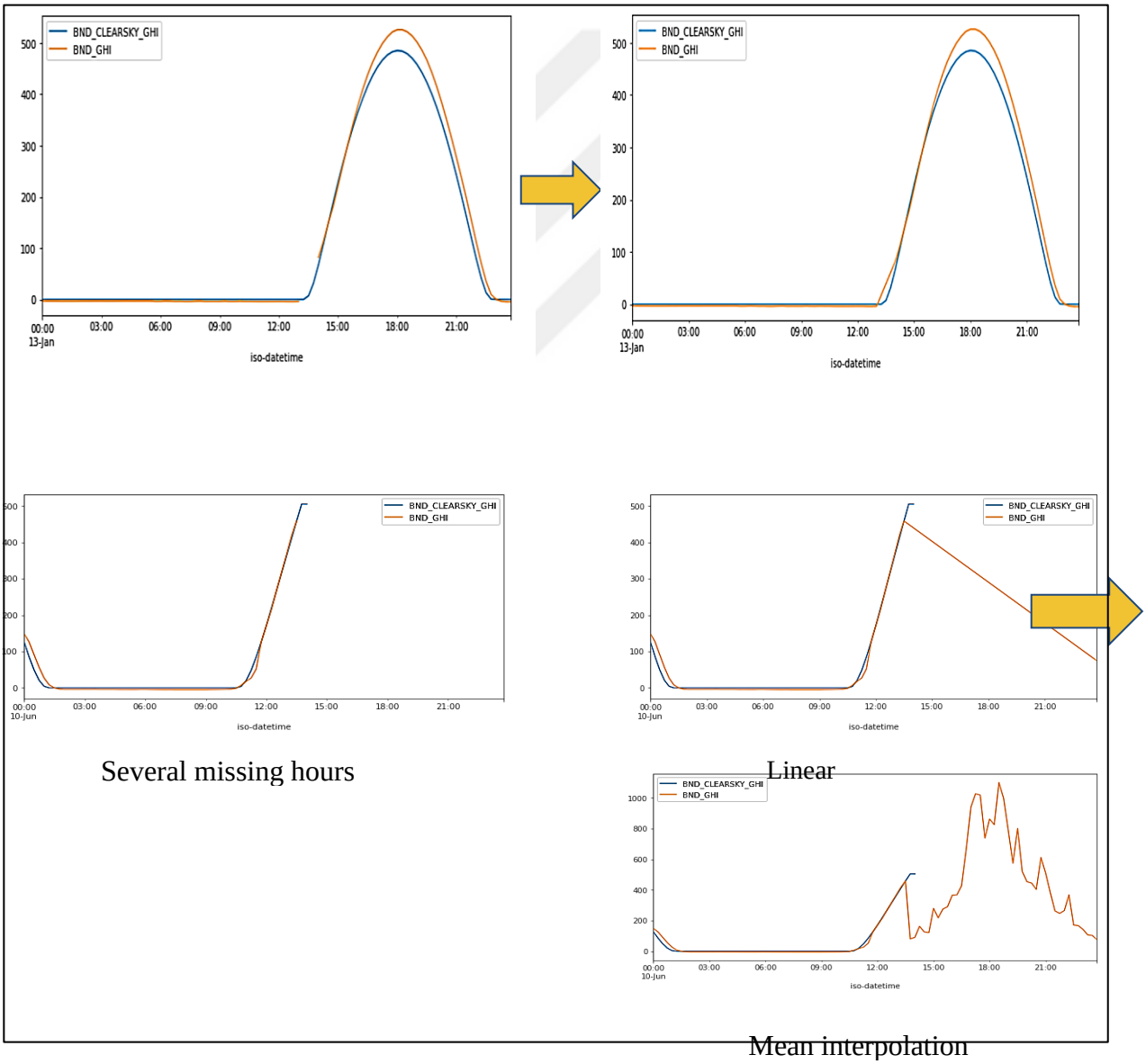

Figure 3.6: Linear interpolation

We interpolated missing GHIs using a mean interpolation based on the previous and next day of valid data.

3.3.6 Dealing with Missing Images

We experienced with interpolating missing images by using a linear interpolation pixels-wise from previous and next available images:

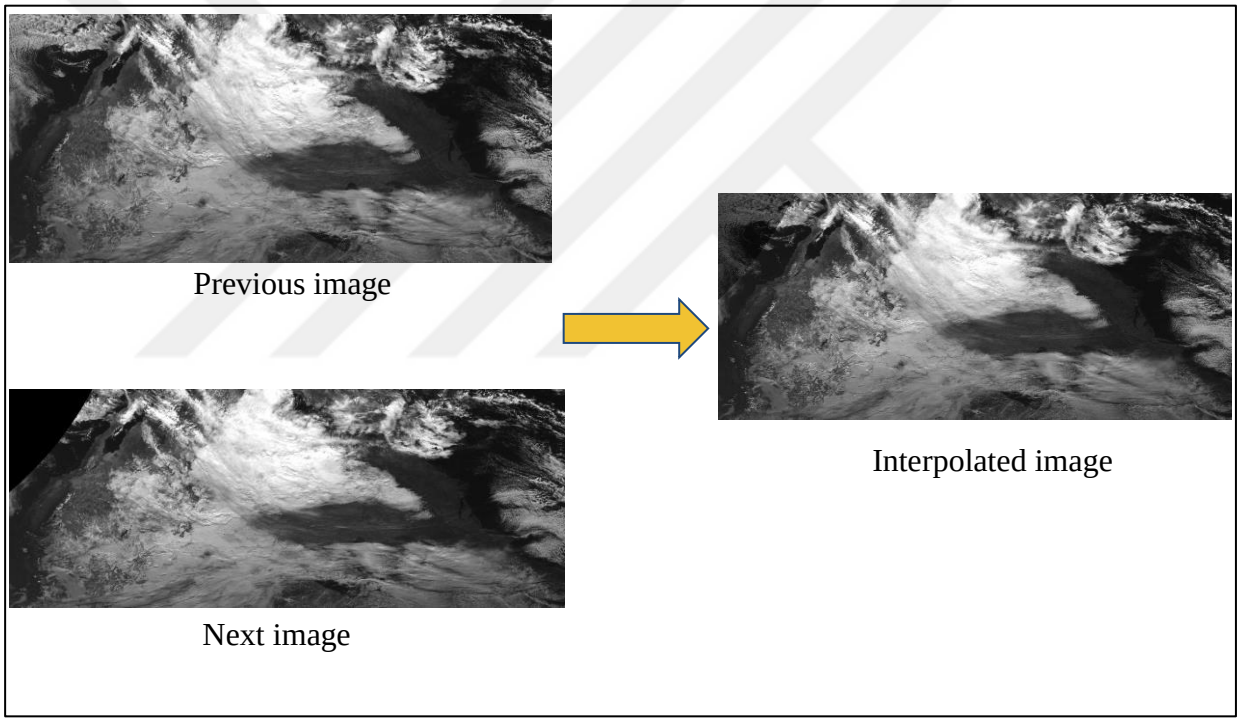

Figure 3.7: The dynamics of atmospheric

In order to capture the dynamics of atmospheric phenomena, we considered a sequence length of past 2 images with a window size of 41. Thus, our input image is of size (82,82,5).

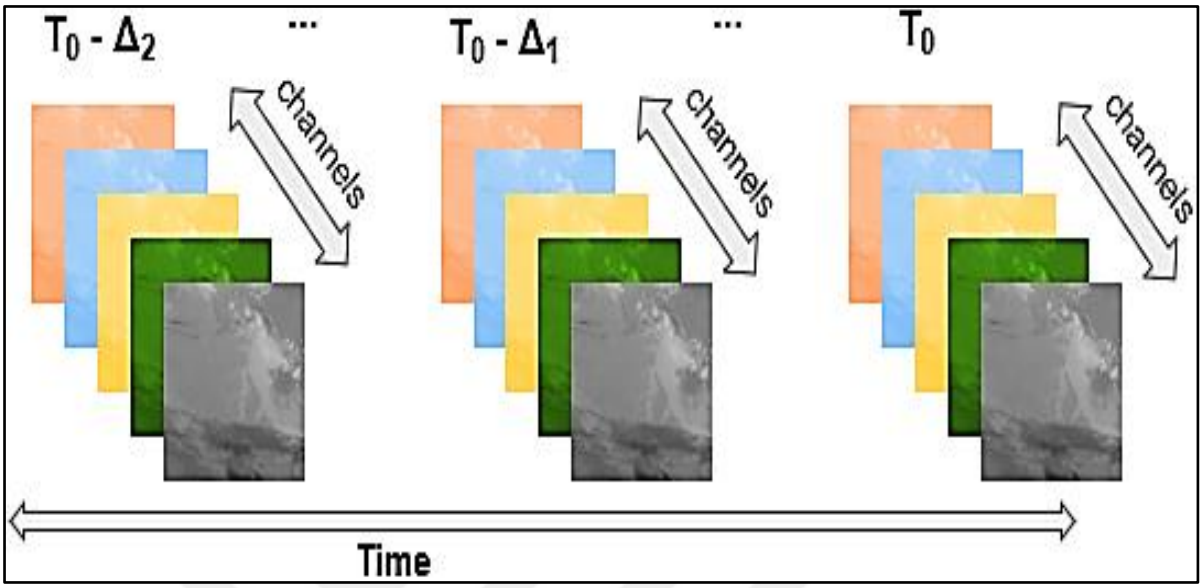

Figure 3.8: Missing in the pre-defined sequence length.

If there was any missing image in the pre-defined sequence length, we replaced those with the image at T_0

3.3.7 Cyclical Solar Data

Another processing step that we did was to encode cyclical continuous features such as the day of the years and the time of the day such that our model knew they were cyclical. We did so by transforming the data in 2D using a sin and cosine transformation.

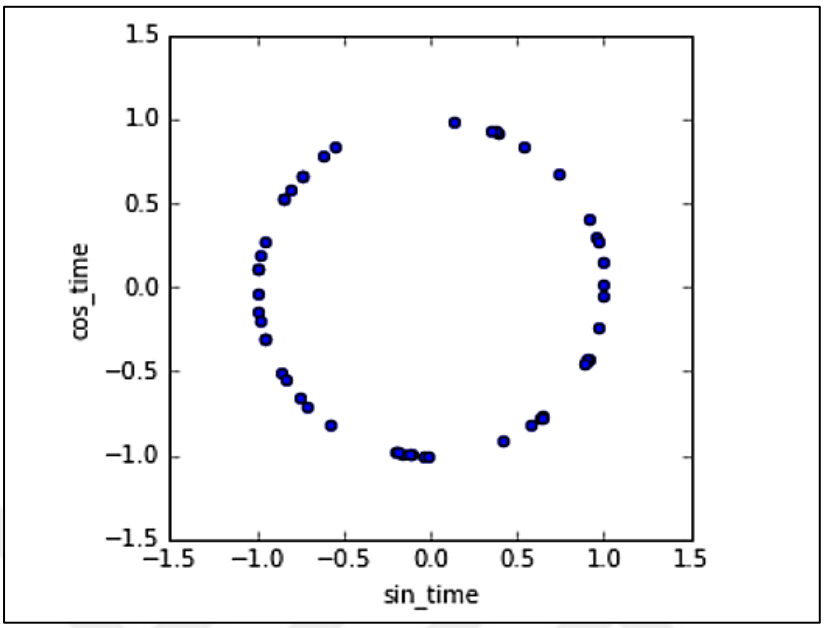

Figure 3.9: Cyclical Solar Data

3.3.8 Model Building

Once we had the cropped images, we wanted to save those images so our data loader could access them. For each $T0$ timestamp alongside its previous sequence of images, we defined a mini-batch of 256 that we saved in a .H5 file in addition to the true GHI, the clearsky GHI and the daytime flag for $T0$,$T1$,$T3$,$T6$. Also, in this .H5 file, we stored the station ID and the true datetime for each station, that converts the global timestamp to its own local timestamp according to the timezone of each station. These additional features were useful for the other models tried, such as the ConvLSTM [20] mentioned in the literature review. Those files were stored in our team directory on Helios server and served as input to our model. Refer to figures 5 and 6 for an illustration of the final model design. Refer to table 3 for a description of the convolutional layers used by the model. Note that after the convolutional layers, we stack normalized clearsky GHI values and time of day and time of year cyclical features with the output of the CNN layers, and pass these inputs on to the dense layers.

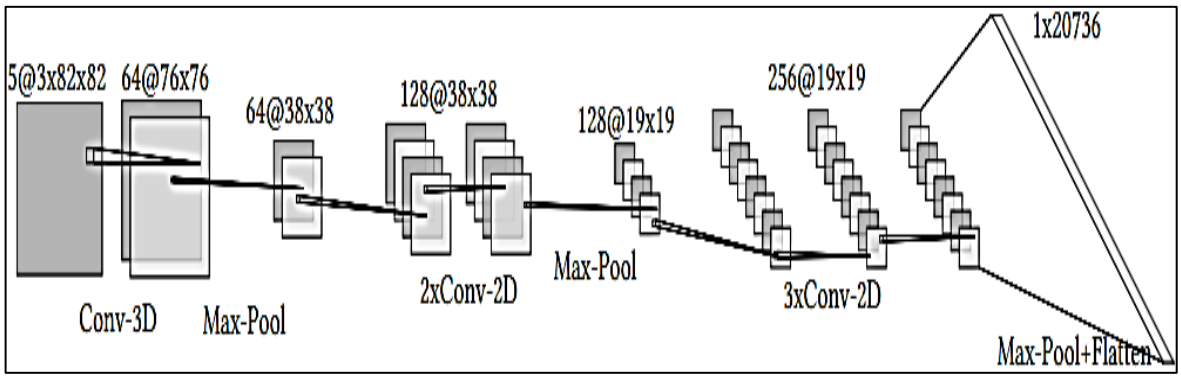

Figure 3.10: 3D-CNN architecture

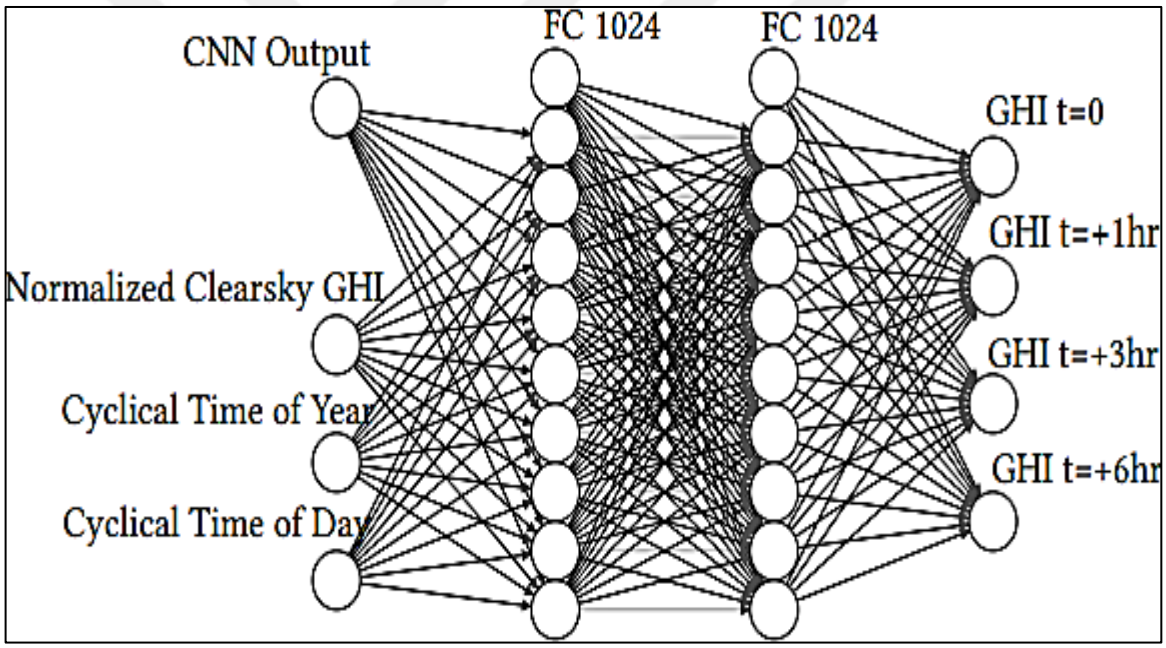

Figure 3.11: Dense layer of the 3D-CNN

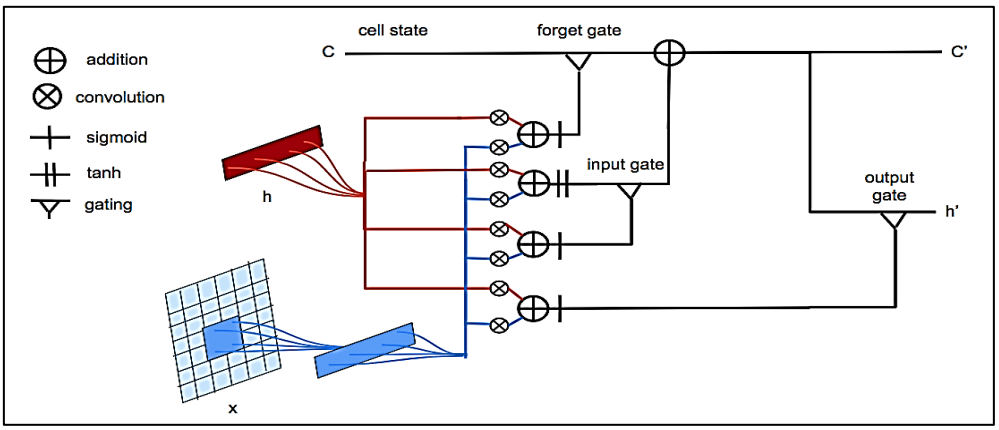

Figure 3.12: Convolutional LSTM

We implement a feed-forward CNN with architecture similar to VGGnet [2] and AlexNet [13]. To summarize, our model consists of:

i- One 3D CNN Layer

ii- 5 2D CNN Layers

iii- 3 max pooling layers

iv- Batch Normalization [21] after every layer

v- RELU activation function after each batch normalization operation

vi- 3 dense layers after CNN portion of model

vii- Sigmoid output activation function

Refer to figures 5 and 6 for an illustration of the final model design. Refer to table 3 for a description of the convolutional layers used by the model. Note that after the convolutional layers, we stack normalized clearsky GHI values and time of day and time of year cyclical features with the output of the CNN layers and pass these inputs on to the dense layers.

The output of the convolutional network is then stacked with additional input data. This data includes the cyclical time of day, cyclical time of year, and the normalized clearsky GHI value.

3.4 PERFORMANCE EVALUATION

To guarantee that the investigation is essentially as intensive as could be expected, the tests should be performed under similar circumstances at every possible opportunity. To try not to imperil the presentation investigation, all tests were regulated on a similar PC. Each test was done multiple times. This was because of the way that a few results were not steady between reproductions. The reproduction time in the two techniques differs marginally relying upon the speed of the registering processes. Since the beginning organization's loads are set aimlessly in the ANN situation, the worth of the conjecture for which the model combines aren't a similar all the time. Subsequently, to gain a precise outcome, we arrived at the midpoint of the ten consequences of each test reiteration.

To obtain the exhibition mistake of each estimate and to all the more likely look at every boundary and situation, different numerical techniques that are regularly utilized in this sort of examination were required:

a. Root Mean Square Error:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(z_t - SI_{real})^2}{N}} \tag{3.2}$$

b. Mean Absolute Error:

$$MAE = \frac{\sum_{t=1}^{N}|z_t - SI_{real}|}{N} \tag{3.3}$$

c. Mean Absolute Percentage Error:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{z_t - SI_{real}}{SI_{real}}\right| \tag{3.4}$$

Where zt is the expected worth of sun powered irradiance, SIreal is the real worth, and N is the absolute number of gauges. Since we know deduced that the conjecture for the nighttime

frame is equivalent to nothing, the calculation of the presentation blunder for this period isn't considered in all tests. At the point when the genuine Sun oriented Irradiance surpasses 10 W/m2, we will assess the mistake. The MAPE is the essential assessment model in this paper. In any case, given likely uncertainties between particular speculations, we applied different models. In different conditions, these three classifications of flaws were deficient to figure out which situation was awesome. Then, at that point, as displayed in the accompanying condition, we determined the connection between gauge mistake and Constancy Model.

$$MAPE_{model} = \frac{MAPE_{pred}}{MAPE_{pers}} \tag{3.5}$$

If the prediction matches the result in the Persistence Model, the MAP Emodel equals 1. The smaller the MAP Emodel, the better the performance, especially when compared to the Persistence Model.

4. RESULT AND DISCUSSION

4.1 DEVELOPMENT ENVIRONMENT

Python is a general-purpose programming language with a high level of abstraction. Python, developed by Guido van Rossum and introduced in 1991, has a design philosophy that prioritizes code readability, including the use of virtual spaces. It gives developments that permit unequivocal programming at a little and huge scope. Python has a unique device and computerized memory board. It helps with more than one programming standard, comprehensive of article situated, basic, useful, and procedural, and has a total general library. Python translators are reachable for the overwhelming majority of running designs. The Jupyter Scratch cushion is an open-source web application that permits you to make and trade records containing stay code, conditions, portrayals, and story text-based material. Data purifying and change, numerical reenactment, authentic appearance, information portrayal, man-made intelligence, and various more applications are models.

4.2 EXPLORATORY DATA ANALYSIS

This fig shows, the data column has Date and Time in a single column which needs to be splitted into Date, Month & Year and Time needs to be spitted into Hours, Minutes & Seconds so as to get some better inferences.

	UNIXTime	Data	Time	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	TimeSunRise	TimeSunSet
0	1475229326	9/29/2016 12:00:00 AM	23:55:26	1.21	48	30.46	59	177.39	5.62	06:13:00	18:13:00
1	1475229023	9/29/2016 12:00:00 AM	23:50:23	1.21	48	30.46	58	176.78	3.37	06:13:00	18:13:00
2	1475228726	9/29/2016 12:00:00 AM	23:45:26	1.23	48	30.46	57	158.75	3.37	06:13:00	18:13:00
3	1475228421	9/29/2016 12:00:00 AM	23:40:21	1.21	48	30.46	60	137.71	3.37	06:13:00	18:13:00
4	1475228124	9/29/2016 12:00:00 AM	23:35:24	1.17	48	30.46	62	104.95	5.62	06:13:00	18:13:00

Figure 4.1: Shape of the Data

4.2.1 Correlation Matrix

A correlation matrix is a table that shows the coefficients of connection between's factors. Every cell in the table addresses the connection between two factors. A connection lattice is utilized to sum up information, as a contribution to a more complex examination, and as an indicative for cutting edge investigations.

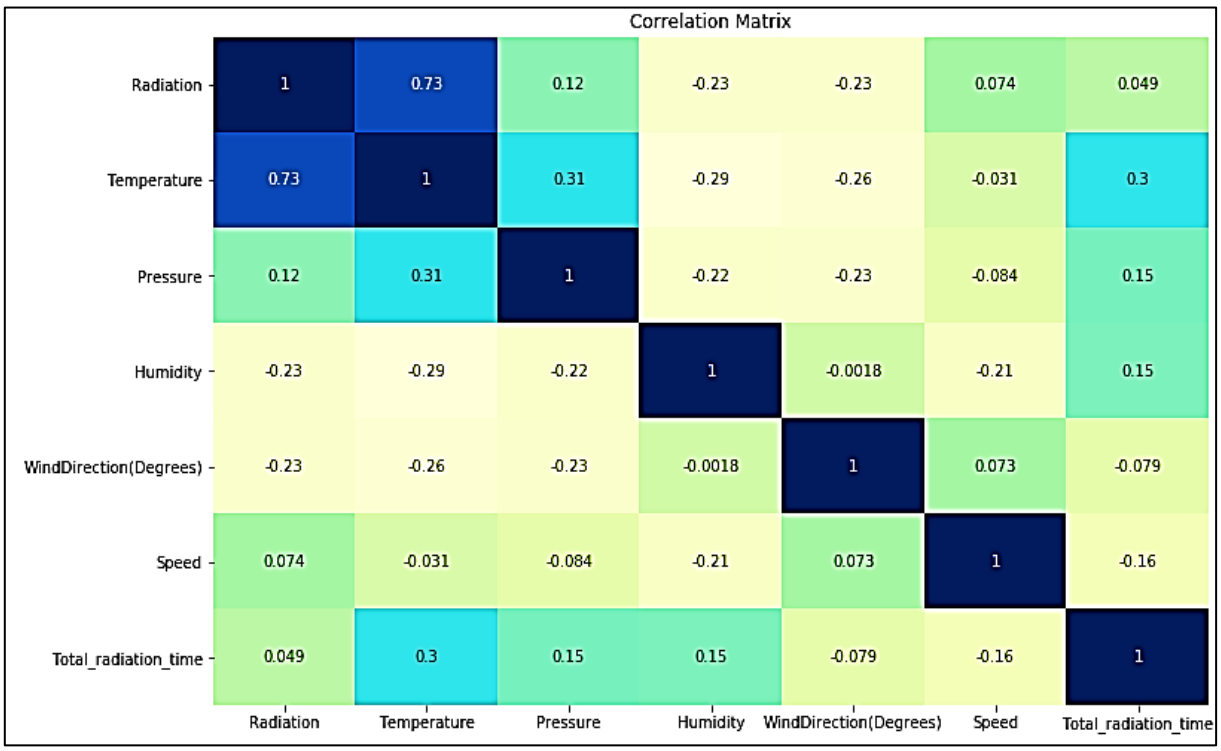

Figure 4.2: Confusion Matrix

There is a high correlation of Radiation with Temperature is 0.73 which means with the increase in the Temperature there will be increase in the Radiation. Apart from it there is no such any major correlation.

4.2.2 Scatterplot

A scatter plot represents the values of two numerical variables with dots. Each dot on the horizontal and vertical axes indicates a single data point's value. Scatter plots are used to investigate the relationships between variables.

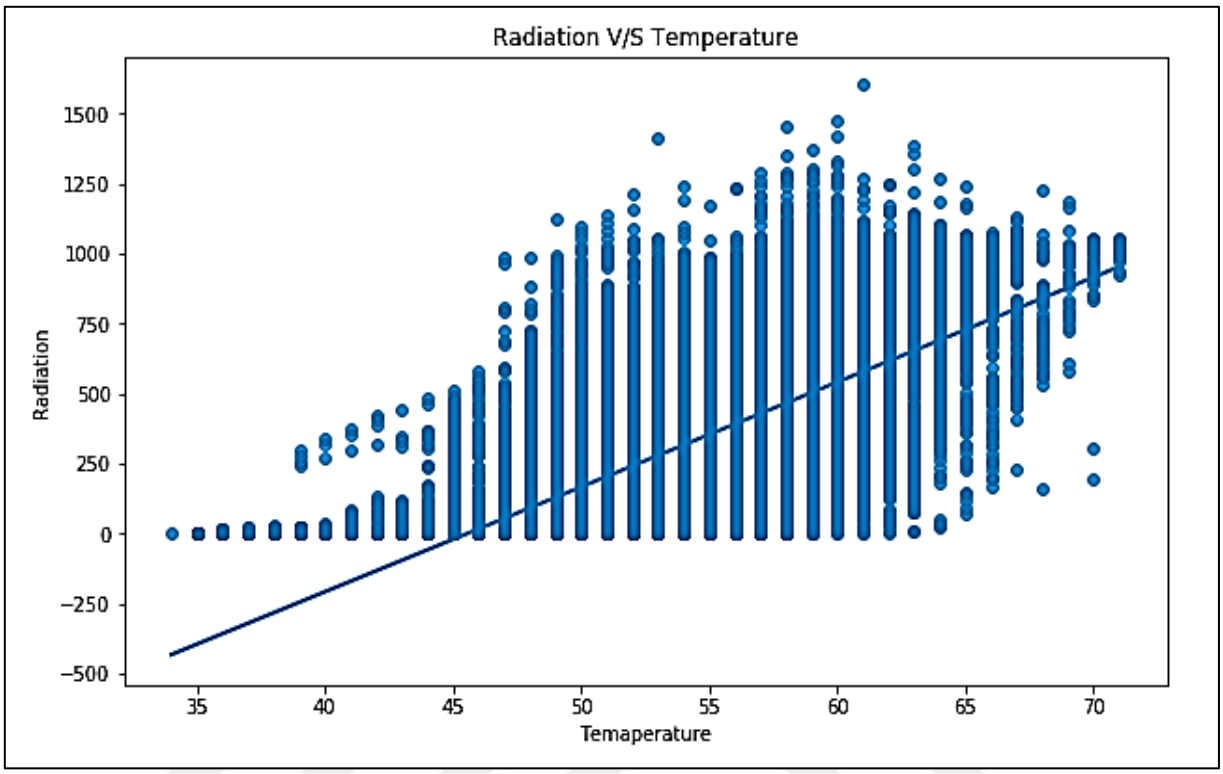

Figure 4.3: Values for two different numeric variables

The scatterplot has a positive Linear Relationship. But on the upper side there are more number of data points as compared to the Lower side of the best fit line. However, the best fit line is showing the positive linear relationship between both the variables.

4.2.3 Outliers

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population.

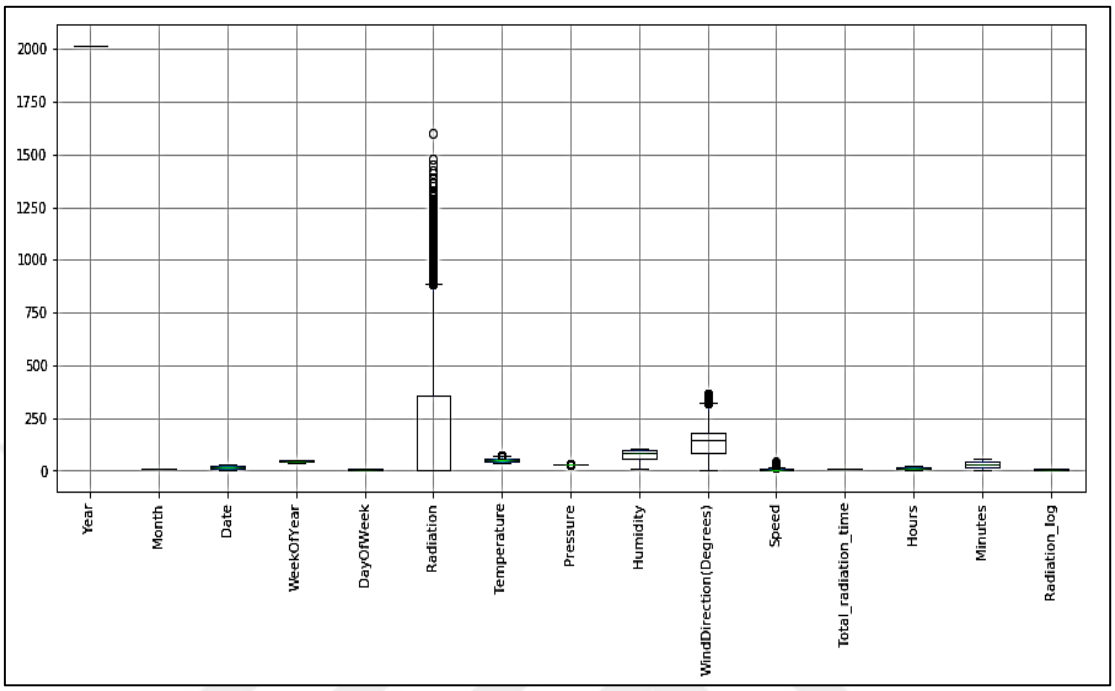

Figure 4.4: Outliers Plot

There are outliers in Radiation, Temperature, Pressure, Wind Direction, and Speed. Since Radiation is our dependent variable therefore, we need to Log transform it & other column needs to be treated with the outlier treatment. The technique we will be using is Winsorization. Winsorization is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers.

By performing the IQR technique, remove the outliers. We were losing 15% of the data. So in order to resolve that issue we have replaced the upper extreme values with upper boundary of the boxplot while we have replaced the lower extreme values with lower boundary of boxplot.

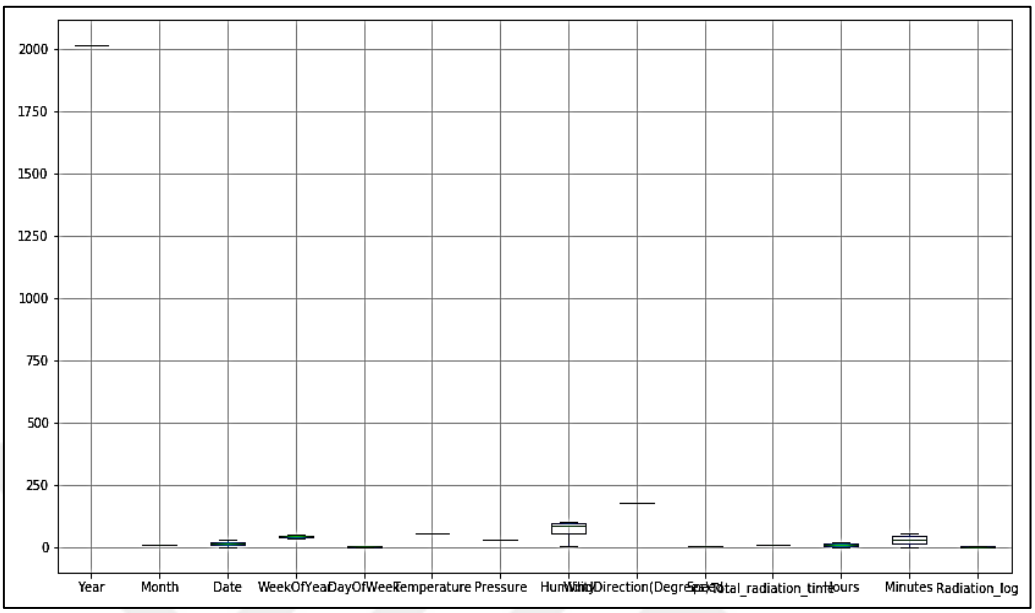

Figure 4.5: Shape of the data after treating the Outliers

Outliers have been removed from the data without losing the original shape.

4.3 CROSS-VALIDATION TECHNIQUES

The model is trained on all data from 2010-2014 and it is validated on data from year 2015 only. The objective of this choice of cross-validation split is to ensure that the data is distributed evenly. The weather over one year is an annual cyclical event, so to ensure an even distribution of data, we aim to select data for training and validation from periods of time that cover this whole cyclical event. Also note that GHIs that occur while 'night_flag' is set are excluded from scoring calculations as to not have the model learn on irrelevant data.

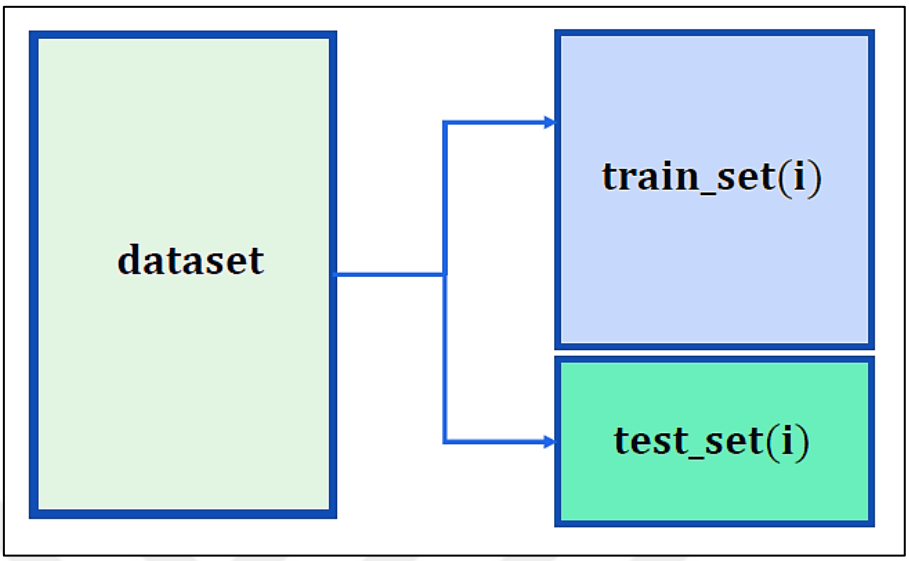

Figure 4.6: Cross-Validation Strategy

Input features:

i. Past 3 cropped images centered around station coordinates
ii. Clearsky value GHIs for $T_{0,}$ $T_{1,}$ $T_{3,}$ T_6
iii. Cyclical time of day
iv. Cyclical time of year

4.4 FINAL HYPERPARAMETERS

After hyperparameter optimization was performed, the following were used in the 'best model': the number of cropped images is set to 3, where the time difference between image captures is 30 minutes. The final batch size is 256 units, whose choice was a tradeoff between performance and memory limits. The learning rate used was (0.001) with the Adam optimizer. Some hyperparameters that were fixed without substantive tuning are the dropout rate (0), the crop window size (82x82) and the $maxk$ GHI. Note that $maxk$ GHI is defined to be the maximum clearsky ratio, so for example, a value of 1.1 would imply a maximum prediction of 110% of the clearsky value at that time. Some experimentation was performed with dropout vs batch normalization, and it was found that better performance was obtained with batch normalization. Additionally, the number of dense nodes varied by multiples of 2. In the end, the best performance was obtained with 1024 nodes in all 3 dense layers.

4.5 RESULT PRESENTATION

Refer to figures 6 and 7 for the training curves which represent the best model's loss and GHI RMSE performance as a function of training epoch. Note that only 3 epochs of training are required for convergence. Note that the small number of epochs required is also a function of the number of training images available, which is large since our training set spans many years and includes data which might share similar characteristics with other random samples from the training set.

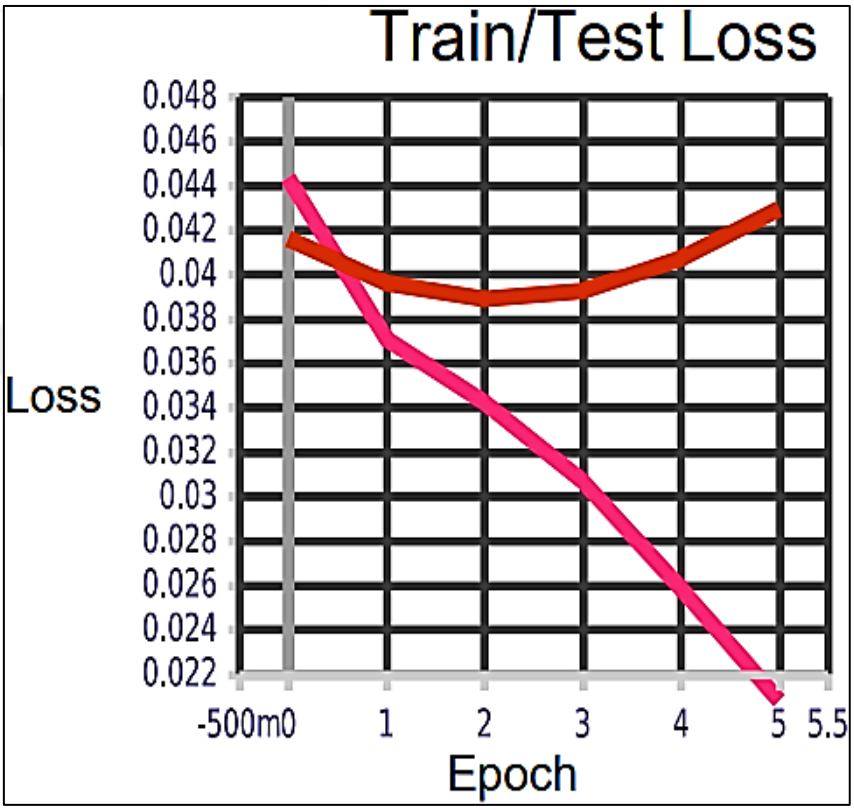

Figure 4.7: Loss over training epochs for the best model

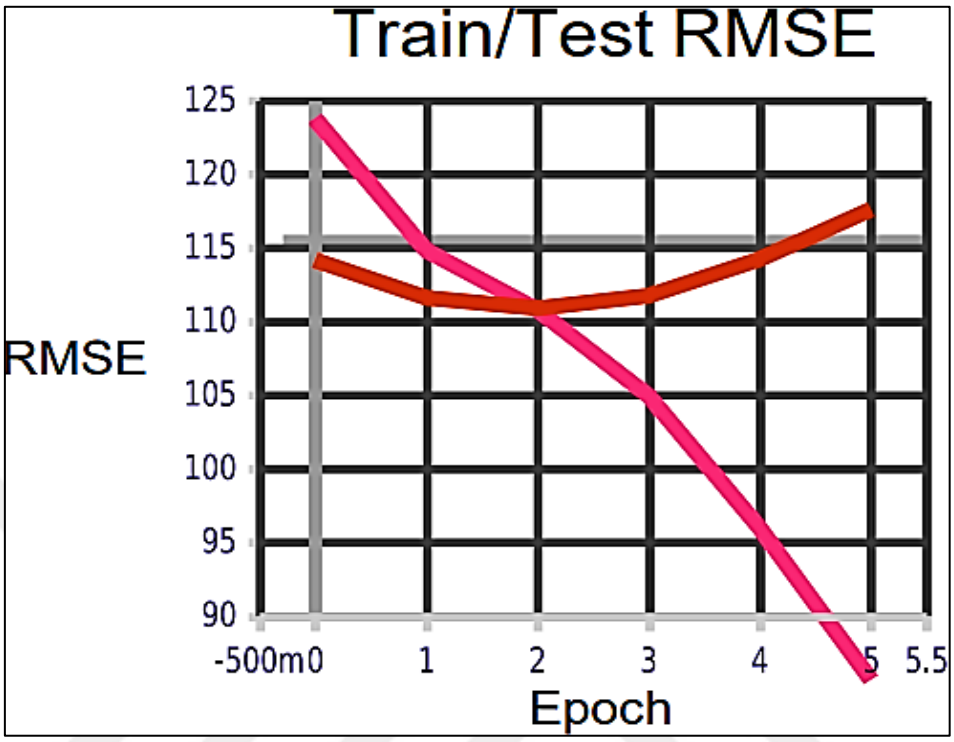

Figure 4.8: RMSE performance over training epochs for the best model

Table 4.1: Performance Summary of Various Models

Model Description	Val RMSE
Best model (ref hyperparmeter section)	110.2
Model with sequence of 5 images	112.2
Dense layers have 2048 Nodes	111.5
Add onehot station IDs	110.9
Clearsky baseline	163.0

While our model beats the clearsky deterministic model, it does not achieve state of the art performance. We suggest a few improvements such as data augmentation, i.e rotating, flipping, and zooming the images. In addition, as stated above, manual hyperparameter search was conducted. Other than typical grid search and random search, Bayesian optimization could be tried to improve the performance of the model. All through this work, it was feasible to more deeply study computerized reasoning procedures, specifically about the ML and DL models and how these models could be applied to sun powered irradiance Forecast.

5. CONCLUSIONS AND FUTURE WORKS

5.1 CONCLUSIONS

As we have seen in the introduction to this thesis, the Solar system sizing requires reliable, comprehensive and long-term solar radiation data. To defeat the absence of long series of dependable sunlight-based information, vital for the advancement and ideal measuring of nearby planet groups, a way to deal with foresee hourly worldwide sun oriented radiation from less expensive meteorological information was introduced in this composition. Ten distinct relationships of five meteorological factors were utilized to foster two sorts of ANN models. In this study, a CNN-based model utilizing GOES-13 satellite imagery and related metadata for nowcasting GHIs prediction was proposed. The main advantage of the 3D-CNN model we propose is its faculty to represent complex spatial and temporal phenomena like the displacement of atmospheric conditions such as clouds. We demonstrate the validity of the model's capability of such representation by comparing its performance with that of the clearsky deterministic model baseline. We were constrained due to tight project deadlines. Given more time, we would have liked to explore many things such as: In the development of our model, as stated in the methodology, we manually tuned hyperparameters. Future work might include performing a more sophisticated hyperparameter search with Orion or using some other parallel hyperparameter searching tool.

5.2 FUTURE WORKS

This study decided the ideal environment variable blend to utilize in a recommended model for globe sun powered radiation (GSR) estimating that was fundamental in structure and utilized made up mind associations. Date, air temperature, and relative wetness were the best mix of environment qualities that fulfilled the objectives of effortlessness, ideal number of neurons, and further developed execution (as assessed by r, RMSE, and MAPE). Regardless, this model beat as of late circulated models that utilized comparative climatic boundaries or a comparable number of neuron network layers.

While this model outflanked distributed models concerning r, RMSE, and MAPE, a few models utilized less neurons or had lower RMSE or MAPE. Despite the fact that daylight and its term over the course of the day were not utilized as climate factors in this review, future work could incorporate examining daylight as a variable and distinguishing a strategy to address consistency challenges that are experienced while utilizing overcast cover as a variable in GSR expectation. Further exploration ought to consider codes and different adjustments to additionally improve on the construction of models for GSR expectation that join multiple factors. Moreover, the impact of season of day and sun radiation point will be thought of. Further investigation into the connection between climate factors and sunlight-based radiation is encouraged. A half breed model for sunlight-based expectation utilizing wavelet brain organizations and brain fluffy organizations could be created from now on.

A model that re-evaluates the times for each region and updates the input masks daily could improve the accuracy of the morning and dusk time regions, as changes in season and obstacles affect dawn and dusk. Finally, our approach provides synthetic solar radiation series for use in the optimal sizing and planning of solar energy systems.

REFERENCES

[1] R. Castro, Uma Introduc¸ao˜ as Energias Renov ` aveis: E ´ olica, Fotovoltaica e Mini-H ´ ´ıdrica, 1st ed. IST Press, 2011.

[2] VON OMPTEDA, K. A. R. I. N. "13 Data Manifestation: Climate Change Data in the Home and on the Body." Design and Science (2023): 271\.

[3] Cox, Anna. "Climate Change Data for New Zealand Schools." (2017).

[4] Kocevska, Katerina, and Bekim Nuhija. "Concession as a Precondition for Relatively Reserved Rights of the Foreigners in the Republic of Macedonia: The Case of Small Hydroelectric Plants." J. Liberty & Int'l Aff. 3 (2017): 55..

[5] Gallo, Michela, et al. "Sustainability in maritime sector: Waste management alternatives evaluated in a circular carbon economy perspective." Resources 9.4 (2020): 41.

[6] Bátora, Jozef. "Reinvigorating the rotating presidency: Slovakia and agenda-setting in the EU's external relations." Global Affairs 3.3 (2017): 251-263.

[7] Carr, Geoffrey. "Sunny Uplands: Alternative energy will no longer be alternative." The Economist 21 (2012).

[8] R. E. P. N. for the 21st Century, "Renewables 2016, Global Status Report," Renewable Energy Policy Network for the 21st Century, Tech. Rep., 2016.

[9] Behrens, Paul, et al. "Environmental, economic, and social impacts of feed-in tariffs: A Portuguese perspective 2000–2010." Applied energy 173 (2016): 309-319.

[10] P. do Conselho de Ministros, "Resoluc¸ao do Conselho de Ministros n ˜ o20/2013," Diario da Rep ´ ubica, ´ 1 aSerie - n ´ o70, April 2013.

[11] D. G. de Energia e Geologia, "Renovaveis: Estat ´ ´ısticas Rapidas," Minist ´ erio da Economia, Tech. ´ Rep. 146, December 2016.

[12] Shi, Jie, et al. "Forecasting power output of photovoltaic systems based on weather classification and support vector machines." IEEE Transactions on Industry Applications 48.3 (2012): 1064-1069.

[13] Monteiro, Claudio, et al. "Short-term power forecasting model for photovoltaic plants based on historical similarity." Energies 6.5 (2013): 2624-2643.

[14] Bigdeli, Nooshin, and Amir Hossein Zandieh. "Solar Irradiance Prediction by a New Forecast Engine Composed Wavelet Packet Transform and Adaptive Neuro-Fuzzy Inference System." IJECCE 4.6 (2013): 1598-1606.

[15] Qing, Xiangyun, and Yugang Niu. "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM." Energy 148 (2018): 461-468.

[16] Kaba, Kazım, et al. "Estimation of daily global solar radiation using deep learning model." Energy 162 (2018): 126-135.

[17] Sargent, Kira M. "Supporting renewable energy: lessons from the deer island treatment plant." (2010).

[18] Gabás Royo, Alberto Eduardo. Solar irradiance forecasting using neural networks. MS thesis. Universitat Politècnica de Catalunya, 2019.

[19] Atwater, M. A_, and J. T. Ball. "A numerical solar radiation model based on standard meteorological observations." Solar Energy 21.3 (1978): 163-170.

[20] Trahi, Fatiha. Prédiction de l'irradiation solaire globale pour la région de Tizi ouzou par les réseaux de neurones artificiels: Application pour le dimensionnement d'une installation photovoltaique pour l'alimentation du laboratoire de recherche LAMPA. Diss. Université Mouloud Mammeri, 2011.

[21] Croci, Lila. Gestion de l'énergie dans un système multi-sources photovoltaïque et éolien avec stockage hybride batteries/supercondensateurs. Diss. Université de Poitiers, 2013.

[22] Motahhir, Saad, Abdelaziz El Ghzizal, and Aziz Derouich. "Modélisation et commande d'un panneau photovoltaïque dans l'environnement PSIM (Modeling and Control of a Photovoltaic Panel in the PSIM Environment)." (2015).

[23] Choi, Rene Y., et al. "Introduction to machine learning, neural networks, and deep learning." Translational Vision Science & Technology 9.2 (2020): 14-14.

[24] Bouguerra, Oussama, and Oussama Benslimane. Solar radiation prediction using machine learning. Diss. Univ M'sila, 2020.

[25] Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.

[26] Bouguerra, Oussama, and Oussama Benslimane. Solar radiation prediction using machine learning. Diss. Univ M'sila, 2020.

[27] Alzahrani, Ahmad, Jonathan W. Kimball, and Cihan Dagli. "Predicting solar irradiance using time series neural networks." Procedia Computer Science 36 (2014): 623-628.

[28] Diaconescu, Eugen. "The use of NARX neural networks to predict chaotic time series." Wseas Transactions on computer research 3.3 (2008): 182-191.

[29] Sfetsos, A., and A. H. Coonick. "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques." Solar Energy 68.2 (2000): 169-178.

[30] Islam, Md Shafiqul, Md Monirul Kabir, and Nafis Kabir. "Artificial neural networks based prediction of insolation on horizontal surfaces for Bangladesh." Procedia Technology 10 (2013): 482-491.

[31] Karlsruhe, U., and L. Prechelt. "Early Stopping| but when." Neural Networks Tricks Trade 1524 (2012): 55-69.

[32] Solmaz, Ozgur, and Muammer Ozgoren. "Prediction of hourly solar radiation in six provinces in turkey by artificial neural networks." Journal of energy engineering 138.4 (2012): 194-204.

[33] Sözen, Adnan, et al. "Use of artificial neural networks for mapping of solar potential in Turkey." Applied Energy 77.3 (2004): 273-286.

[34] Jiang, Yingni. "Computation of monthly mean daily global solar radiation in China using artificial neural networks and comparison with other empirical models." Energy 34.9 (2009): 1276-1283.

[35] Pandey, C. K., and A. K. Katiyar. "Solar radiation: Models and measurement techniques." Journal of Energy 2013 (2013).

[36] Trnka, Miroslav, et al. "Global solar radiation in Central European lowlands estimated by various empirical formulae." Agricultural and Forest Meteorology 131.1-2 (2005): 54-76.

[37] Bahel, V., H. Bakhsh, and Ra Srinivasan. "A correlation for estimation of global solar radiation." Energy 12.2 (1987): 131-135.

[38] Bristow, Keith L., and Gaylon S. Campbell. "On the relationship between incoming solar radiation and daily maximum and minimum temperature." Agricultural and forest meteorology 31.2 (1984): 159-166.

[39] Allen, Richard G. "Self-calibrating method for estimating solar radiation from air temperature." Journal of Hydrologic engineering 2.2 (1997): 56-67.

[40] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).

[41] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[42] Lai, Guokun, et al. "Modeling long-and short-term temporal patterns with deep neural networks (2017)." arXiv preprint arXiv:1703.07015 (2018).

[43] Bertrand, Cédric, Gilles Vanderveken, and Michel Journée. "Evaluation of decomposition models of various complexity to estimate the direct solar irradiance over Belgium." Renewable energy 74 (2015): 618-626.

[44] Hove, T., E. Manyumbu, and G. Rukweza. "Developing an improved global solar radiation map for Zimbabwe through correlating long-term ground-and satellite-based monthly clearness index values." Renewable energy 63 (2014): 687-697.

[45] Palmer, Diane, et al. "Satellite or ground-based measurements for production of site specific hourly irradiance data: Which is most accurate and where?." Solar Energy 165 (2018): 240-255.

[46] Besharat, Fariba, Ali A. Dehghan, and Ahmad R. Faghih. "Empirical models for estimating global solar radiation: A review and case study." Renewable and Sustainable Energy Reviews 21 (2013): 798-821.

[47] Despotovic, Milan, et al. "Review and statistical analysis of different global solar radiation sunshine models." Renewable and Sustainable Energy Reviews 52 (2015): 1869-1880.