



**REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY
UNIVERSITY**

**GRADUATE SCHOOL OF
DEPARTMENT OF ELECTRIC AND ELECTRONIC ENGINEERING**

**DIGITIZING ELECTRICITY GENERATION IN SOLAR PANEL
FARMS WITH DIGITAL TWIN AND OPERATING WITH HIGH
EFFICIENCY**

**TOLGA YALÇIN
MASTER OF SCIENCE**



**REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY
UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRIC AND ELECTRONIC ENGINEERING**

**DIGITIZING ELECTRICITY GENERATION IN SOLAR PANEL
FARMS WITH DIGITAL TWIN AND OPERATING WITH HIGH
EFFICIENCY**

**TOLGA YALÇIN
MASTER OF SCIENCE**

**SUPERVISOR
Assoc. Prof. Dr. TUĞÇE DEMİRDELEN**

ADANA 2023

DECLARATION OF CONFORMITY

In this thesis study, which was prepared following the thesis writing rules of Adana Alparslan Türkeş Science and Technology University Institute of Graduate School, I declare that I provide all the information, documents, evaluations and results in accordance with scientific ethics and moral codes without resorting to any means or assistance that would be contrary to scientific ethics and traditions. I also declare that I refer to all of the articles I used in this study with appropriate references and accept all moral and legal consequences if a situation is found contrary to my statement regarding my work.

19/06/2023

[Signature]

Tolga YALÇIN

ABSTRACT

DIGITIZING THE SOLAR PANEL PRODUCTION SYSTEM WITH DIGITAL TWIN AND OPERATING IT WITH HIGH EFFICIENCY

Tolga YALÇIN

Department of Electric and Electronic Engineering

Supervisor: Assoc. Prof. Dr. Tuğçe DEMİRDELEN

June 2023, 84 pages

Renewable energy sources, particularly solar energy, have witnessed significant growth in popularity and production capacity. Over 1000 GW of photovoltaic panels have been installed and operational worldwide by the end of 2022. However, ensuring high-energy production efficiency, preventing, and detecting errors in the system have become increasingly urgent. The non-linear nature and unpredictability of solar power production make predictability and instantaneous change in production challenging. Therefore, this thesis aims to enhance the traceability of solar power plants by developing a digital twin and creating a platform utilizing the necessary tools to optimize the system's performance.

Initially designed in Matlab, the solar power plant was moved to a simulation environment to evaluate its operating characteristics and refine its design. The data collected were used to train machine learning algorithms, which are employed to monitor the system, predict future scenarios, and notify users when an unusual event or power loss occurs. Additionally, an algorithm was created to address contamination issues and optimize panel cleaning. The integration of these methods and programs, along with flexible interconnection, provides enhanced data visualization, improving traceability.

It is hoped that this thesis will contribute to researchers and plant owners who want to examine solar power plant behavior and increase the efficiency and traceability of solar power plants.

Keywords: digital twin, solar plants, machine learning, O&M systems

ÖZET

GÜNEŞ PANELİ ÜRETİM SİSTEMİNİN DİJİTAL İKİZ İLE DİJİTALİZASYONU VE YÜKSEK VERİMLİLİKLE ÇALIŞTIRILMASI

Tolga YALÇIN

Elektrik ve Elektronik Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Tuğçe DEMİRDELEN

Haziran 2023, 84 sayfa

Yenilenebilir enerji kaynakları, özellikle güneş enerjisi, popüleritesinde ve üretim kapasitesinde önemli bir artışa tanık oldu. 2022'nin sonuna kadar dünya çapında 1000 GW'ın üzerinde fotovoltaik panel kuruldu ve faaliyete geçti. Ancak, yüksek enerji üretim verimliliği sağlamak, sistemdeki hataları önlemek ve tespit etmek giderek daha acil hale geldi. Güneş enerjisi üretiminin doğrusal olmayan doğası ve öngörülemezliği, üretimde öngörülebilirliği ve anlık değişimi zorlaştırır. Bu nedenle, bu tez, dijital bir ikiz geliştirerek ve sistemin performansını optimize etmek için gerekli araçları kullanan bir platform oluşturarak güneş enerjisi santrallerinin izlenebilirliğini artırmayı amaçlamaktadır.

Başlangıçta Matlab'da tasarlanan güneş enerjisi santrali, çalışma karakteristiklerini değerlendirmek ve tasarımını iyileştirmek için bir simülasyon ortamına taşınmıştır. Toplanan veriler, sistemi izlemek, gelecekteki senaryoları tahmin etmek ve olağandışı bir olay veya güç kaybı meydana geldiğinde kullanıcıları bilgilendirmek için kullanılan makine öğrenimi algoritmalarını eğitmek için kullanıldı. Ek olarak, kirlenme sorunlarını gidermek ve panel temizliğini optimize etmek için bir algoritma oluşturuldu. Esnek ara bağlantı ile birlikte bu yöntemlerin ve programların entegrasyonu, izlenebilirliği iyileştirerek gelişmiş veri görselleştirme sağlar.

Bu tezin, güneş enerjisi santrali davranışını incelemek ve güneş enerjisi santrallerinin etkinliğini ve izlenebilirliğini artırmak isteyen araştırmacılara ve tesis sahiplerine katkı sağlayacağı umulmaktadır.

Anahtar Kelimeler: dijital ikiz, güneş santralleri, makine öğrenimi, O&M sistemleri



To my family and Gazi Mustafa Kemal ATATÜRK...

ACKNOWLEDGEMENTS

I would like to thank my adviser Assoc. Prof. Dr. Tuğçe DEMİRDELEN who has always been exemplary of her knowledge, business approach, skills and a broad perspective. Also, who believed in me from the very beginning throughout my studies, I always admired her patience, vision, and knowledge, and learned something different in each of our conversations.

Many thanks to my esteemed research assistant Burak ESENBOGA for his valuable criticism, opinions, suggestions, and contributions regarding my thesis.

I would like to thank my boss Jose Luis DOMINGUEZ, who gave me a chance as an intern and supported me every day since then, and made me work on this subject, and my dear colleague Pol Paradell SOLA, who supported me from the first day and helped me whenever I had difficulties.

I would like to thank my colleagues who answered my questions in the preparation of the content of the thesis and in the application part.

Finally, I would like to thank my family who have always been with me on this journey and supported me even when I lost my motivation or my way and had been a light in my life.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	viii
LIST OF FIGURES	x
LIST OF TABLES	xiii
NOMENCLATURE	xiv
1. INTRODUCTION.....	1
2. LITERATURE REVIEW	12
3. MATERIALS AND METHODS	21
3.1 Matlab/Simulink	21
3.1.1 Solar cells	24
3.1.2 Conventional perturb and observe method	30
3.1.3 DC/DC converters	31
3.1.4 Solar pv-grid system	34
3.3 Machine Learning.....	36
3.2.1 ElasticNet regression	38
3.2.2 Multiple linear regression	40
3.2.3 Bayesian regression.....	41
3.2.4 Polynomial regression.....	43
3.2.5 Random forest regression.....	44
3.2.6 Deep neural network regression	47
3.3 Docker Desktop.....	52
3.3.1 InfluxDB.....	54
3.3.2 Redis cache	55
3.3.3 Grafana	56
3.3.4 FastAPI	56

4. RESULTS AND DISCUSSIONS	58
4.1 Pv panel part	59
4.2 DC/DC converter part.....	66
4.3 Power grid part.....	71
4.4 Cleaning optimizer	75
4.5 Monitoring and evaluation part.....	80
5. CONCLUSIONS	82
6. RECOMMENDATIONS.....	84
REFERENCES.....	85
ANNEXES.....	93

LIST OF FIGURES

ss

Figure 1. 1 Renewable capacity all around the world (International Renewable Energy Agency, 2021).....	3
Figure 1. 2 Types and percentage of energy used to produce electricity in the world.....	4
Figure 1. 3 Cost estimation of electricity generation from solar and wind energy in the future	5
Figure 1. 4 Digital twin data flow and processing recommended in this study	9
Figure 1. 5 Virtual representation of the solar panel model in proposed method	10
Figure 3. 1 Simulation Model of the Solar Panel System used.....	23
Figure 3. 2 The results obtained in the simulation can be transferred as a table thanks to this diagram.....	23
Figure 3. 3 Circuit diagram of solar panel (Bellia, Youcef, & Fatima, 2014)	25
Figure 3. 4 Flowchart of the conventional P&O algorithm.....	31
Figure 3. 5 Block Diagram of a DC-DC Converter	32
Figure 3. 6 DC-to-DC Boost Converter basic circuit diagram.....	33
Figure 3. 7 Ideal Transformer Schema	35
Figure 3. 8 Architecture of random forest algorithm	46
Figure 3. 9 Generic architecture of a Multi-Layer Perceptron.	48
Figure 3. 10 The Docker Desktop schematic designed for this research	52
Figure 3. 11 The Docker Desktop schematic designed for this research	53
Figure 3. 12 The Web interface that the user will use to reach the digital twin.....	57
Figure 4. 1 It shows the whole system by dividing it into 3 partitions and states that it has been examined in this way here.	58
Figure 4. 2 It shows the input variables used to train Machine Learning algorithms and the outputs they are intended to predict.	58
Figure 4. 3 In this figure, you can see the correlation matrix of the variables between each other.....	59
Figure 4. 4 In this figure, you can examine the connection between the produced power from the panel and the variables effect this production of panel. (Each dot shows the value of the energy produced on the panel and the other variable being measured at that moment)	62

Figure 4. 5 The Comparison between the actual production results with the predictions made by the ML models a) Prediction performance of DNN Regression b) Prediction performance of Random Forest Regression	65
Figure 4. 6 Expression in Watts of the difference between the estimates made by ML models and the actual production values a) Error performance of DNN Regression b) Error performance of Random Forest Regression	65
Figure 4. 7 Simulation of DC/DC Boost Converter using Controlled Voltage Source and Signal Generator as Power Source	69
Figure 4. 8 Comparison between the estimates made by DNN Regression of the output voltage/current of the DC/DC converter and the actual results. a. the output voltage of DC/DC converter b. the output current of DC/DC converter.	69
Figure 4. 9 Expression of the difference between the predictions made by DNN Regression and the actual values in terms of voltage/Current a. Difference in voltage values b. Difference in current values.	70
Figure 4. 10 The results obtained from the data used for testing purposes and has not been tampered with a) It shows the estimation and the actual KiloWatts (KW) values on the same graph b.) It shows the difference between the estimated and actual Watts values in terms of KiloWatts (KW) c.) It shows error rate between the estimated and actual Watts values in terms of KiloWatts (KW) d.) It shows error values in terms of mathematical calculations.	72
Figure 4. 11 The results obtained from the system connected to the 100-ohm (resistance, capacitor and inductor) connected in series to the grid part in the simulation to be used for testing purposes. a) It shows the estimation and the actual KiloWatts (KW) values on the same graph b.) It shows the difference between the estimated and actual Watts values in terms of KiloWatts (KW) c.) It shows error rate between the estimated and actual Watts values in terms of KiloWatts (KW) d.) It shows error values in terms of mathematical calculations.....	73
Figure 4. 12 The results obtained from the system connected to the 1000-ohm (resistance, capacitor and inductor) connected in series to the grid part in the simulation to be used for testing purposes. a) It shows the estimation and the actual KiloWatts (KW) values on the same graph b.) It shows the difference between the estimated and actual Watts values in terms of KiloWatts (KW) c.) It shows error rate between the estimated and actual Watts values in terms of KiloWatts (KW) d.) It shows error values in terms of mathematical calculations.....	74
Figure 4. 13 Cleaning optimizer diagram designed for this project.	76

Figure 4. 14 Soiling rate slope.....	78
Figure 4. 15 Dashboard in the grafana show the cleaning parameters and results.....	79
Figure 4. 16 Dashboard in grafana with model input data with grid part	81
Figure 4. 17 Examination of a detailed part of the model input data with the energy results of the grid section	81
Figure 5. 1 Digital Twin application prepared for this research	82



LIST OF TABLES

Table 1.1 Primary energy is defined in this assessment as economically traded fuels, including current renewables used to produce power. Energy from all non-fossil power production sources are compensated for on an input-equivalent level. (BP Statistical Review of World Energy,2021)	2
Table 3. 1 The data we obtained as a result of the simulations	24
Table 4. 1 The names of the models used and their error values	63
Table 4. 2 The names of the models used and their error values	68
Table 4. 3 The names of the models used and their error values	71

NOMENCLATURE

PV	: Photovoltaic
DC	: Direct Current
AC	: Alternating Current
MAE	: Mean Absolute Error
MSE	: Mean Square Error
RMAE	: Root Mean Absolute Error
RF	: Random Forrest
DNN	: Deep Neural Network
KW	: Kilowatt
GW	: Gigawatt
DT	: Digital Twin
IoT	: Internet of Things
IEC	: Electrotechnical Commission
STC	: Standard Test Conditions
MIMO	: Multiple-Input Multiple-Output
AI	: Artificial Intelligence
MLP	: Multi Layer Perceptron
CO ₂	: Carbon Dioxide
PWM	: Pulse-Width Modulation
EMF	: Electromotive Force
UPS	: Uninterruptible Power Supply
SRD	: Soiling Rate Differential

1. INTRODUCTION

Giving a clear definition of energy is challenging. In its most basic form, energy allows everything that happens around us to happen. In other terms, energy is a material's ability to do labor. Many aspects of our existence rely on energy. It is the most important component required for various activities such as heating, transportation, manufacturing, lighting, health, technology, and communication, among others. The amount of energy a country has determines its welfare, competitiveness, and development level. (Global Market Outlook for Solar Power, 2021)

With the advancement of science and technology, the global need for energy is growing by the day. To satisfy the country's needs, it has become unavoidable to look for alternative energy sources, to continuously monitor technological and economic advancements, and to maintain and carry out energy-related developments. The supply and demand problem are the most serious issue in the energy industry. The world's population is continually increasing, and it presently stands at 7.8 billion people. All of these individuals will require energy, increasing global energy consumption. Furthermore, energy consumption per capita is proportional to a country's level of living. (Global Market Outlook for Solar Power, 2021)

The utilization and recyclability of energy resources are categorized. They are classified as renewable energy or non-renewable energy resources based on their consumption. On the other hand, they are classified as main or secondary energy sources based on their convertibility. Non-renewable energy sources are energy sources that are unchanging, finite, and discontinuous in nature. This energy source may be exemplified by fossil fuels such as oil, natural gas, and coal. For years, chemical changes of animal and plant wastes have resulted in the production of fossil fuels. The greatest difficulty that humanity has is that the energy generated is highly reliant on fossil fuels. In nature, fossil fuels are used quicker than they are generated by the photosynthetic process. Oil and gas supplies are fast depleting and extracting the remaining hydrocarbons is becoming more complex technologically. (When Solar Policy Went Digital Task Force Report, n.d.) Renewable energy resources, on the other hand, are resources that can be replenished throughout time and are available for a long period. Examples

of this energy source include solar, wind, geothermal, biomass, and hydropower. (Koç et al. n.d.)

Table 1.1 Primary energy is defined in this assessment as economically traded fuels, including current renewables used to produce power. Energy from all non-fossil power production sources are compensated for on an input-equivalent level. (BP Statistical Review of World Energy,2021)

Exajoules	Oil	Natural gas	Coal	Nuclear energy	Hydro electricity	Renewables	Total	Growth per	Rate annum
								2020	2009-19
Total World	174.20	137.62	151.42	23.98	38.16	31.71	557.10	-4.5%	1.9%
OECD	78.5	63.28	27.46	16.67	13.14	18.04	217.11	-7.7%	0.4%
Non-OECD	95.67	74.34	123.96	7.31	25.02	13.67	339.99	-2.3%	3.1%
European Union	20.03	13.68	5.91	6.11	3.04	6.97	55.74	-8.5%	-0.3%

In 2020, closures were experienced all over the world due to the coronavirus, and the effects of this should be easily seen in the energy sector, primary energy usage fell by 4.5 percent, the first drop in energy consumption since 2009. This decrease was generally driven by oil (-9.7 percent), this means that the decrease in oil corresponds to three-quarters of the overall decrease. Except for renewables (+9.7 percent) and hydro (+1.0 percent), use of all fuels fell. Consumption declined in all areas, with the greatest decrease in Europe (-7.8 percent) and North America (-8.0 percent). The Asia-Pacific region saw the smallest reduction (-1.6 percent) because of the growth in China (+2.1 percent) which is the only country in the top 20 economies where energy needs increase in 2020. The fall in consumption in the other areas varied from -7.8 percent in Central South America also -3.1 percent in the Middle East. Still oil continues to account for the lion's stock of the energy combination (31.2 percent). In addition, coal keeping it is second place in 2020, it is sharing 27.2 percent of total primary energy consumption, increase around 27.1 percent the year before. Renewable energy sources and natural gas both reached new highs of 5.7 percent and 24.7 percent, respectively. Renewable energy sources have already surpassed nuclear energy that accounts for just 4.3 percent of the energy stock. In 2020, hydropower area increased by 0.4 points to 6.9 percent, first time increasing since 2014. (Bp Energy Outlook, 2020)

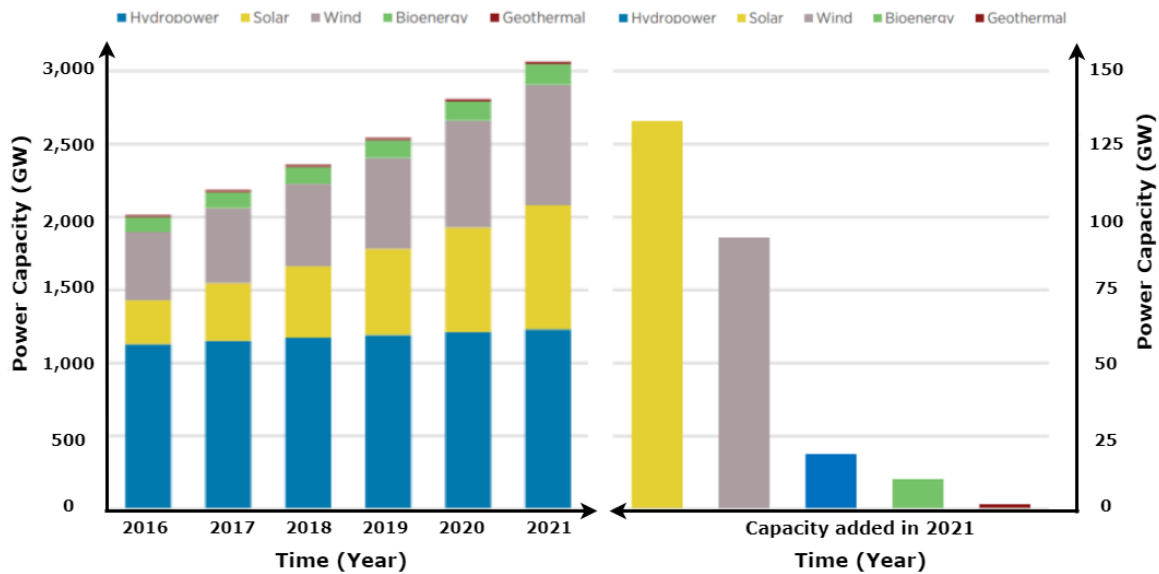


Figure 1.1 Renewable capacity all around the world (International Renewable Energy Agency, 2021)

Global renewable power capacity will reach 3064 GW by the end of 2021. With a capacity of 1 230 GW, hydropower accounted for the biggest percentage of the global total. With capabilities of 849 GW and 825 GW, solar and wind energy provided for an equal part of the remaining. Other renewables included 143 GW of bioenergy, 16 GW of geothermal and 524 MW of marine energy. Renewable energy capacity rose by 257 GW (+9.1%) in 2021. Solar energy continues to lead capacity development, with a rise of 133 GW (+19%) with new capacity added in all major geographical areas in recent years; total worldwide solar capacity has now surpassed wind energy capacity. Wind energy continues to increase by 93 GW (+13%). Hydropower capacity rose by 19 GW (+2%) and bioenergy capacity increased by 10 GW (+8%). Geothermal energy rose by 1.6 GW. (IRENA, 2021)

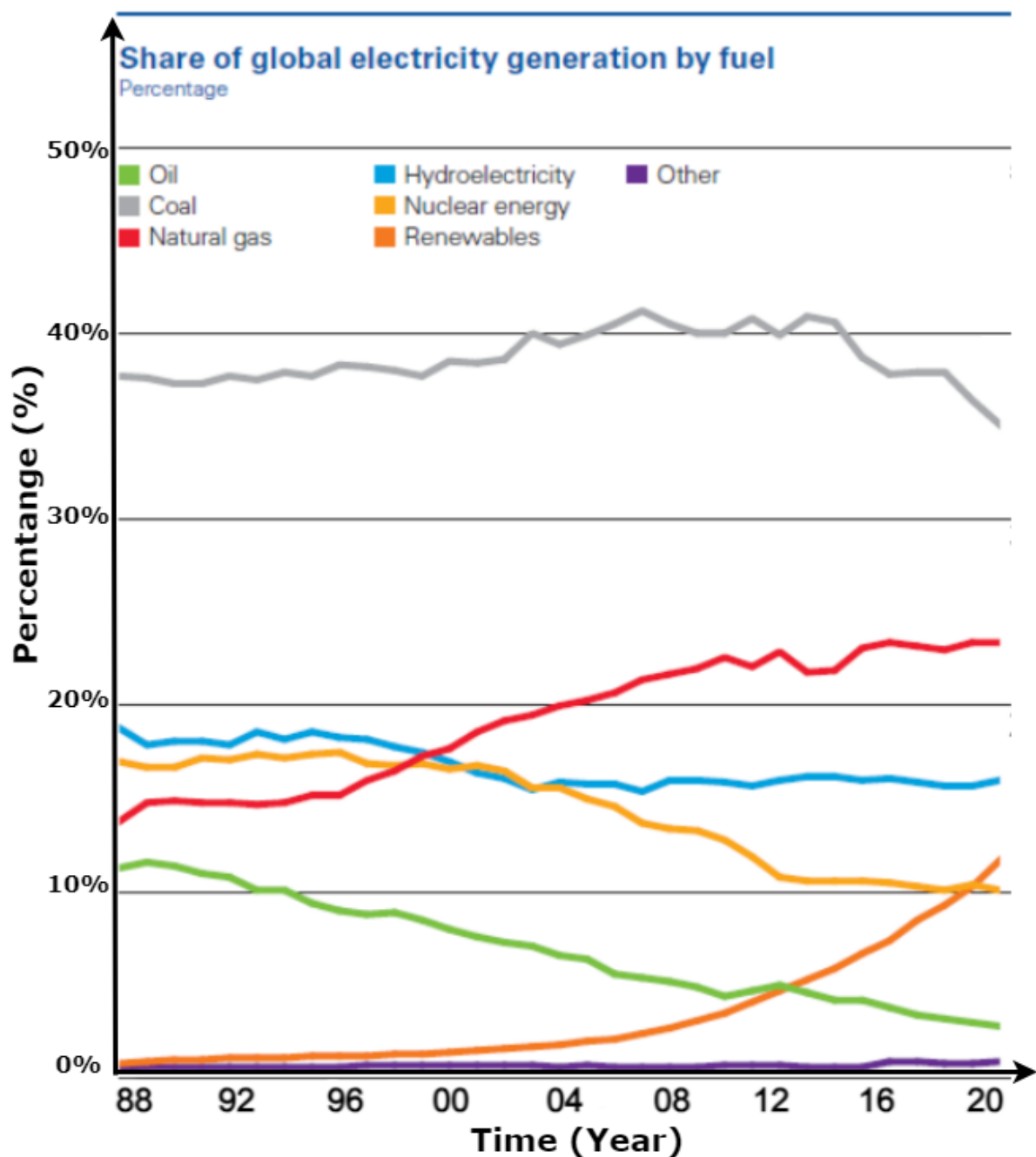


Figure 1. 2 Types and percentage of energy used to produce electricity in the world(BP Statistical Review of World Energy,2021)

Coal remains the leading fuel for electricity production globally (Figure 1.2), however, it shares in the general stock decreased by 1.3 points to 35.1 percent in 2021, the lowest level in the BP data period. Renewable energies hit a new record that is higher than 11.7 percent from last year; if look at the combined proportion of electricity, (35.1 percent) made by gas-fired and renewables surpassing the electricity made by the coal, which is happening first time on the world. Europe's renewable energy contribution in power generation has surpassed nuclear

energy, making Europe the first area where renewables constitute the major source of power generation. (BP Statistical Review of World Energy,2021)

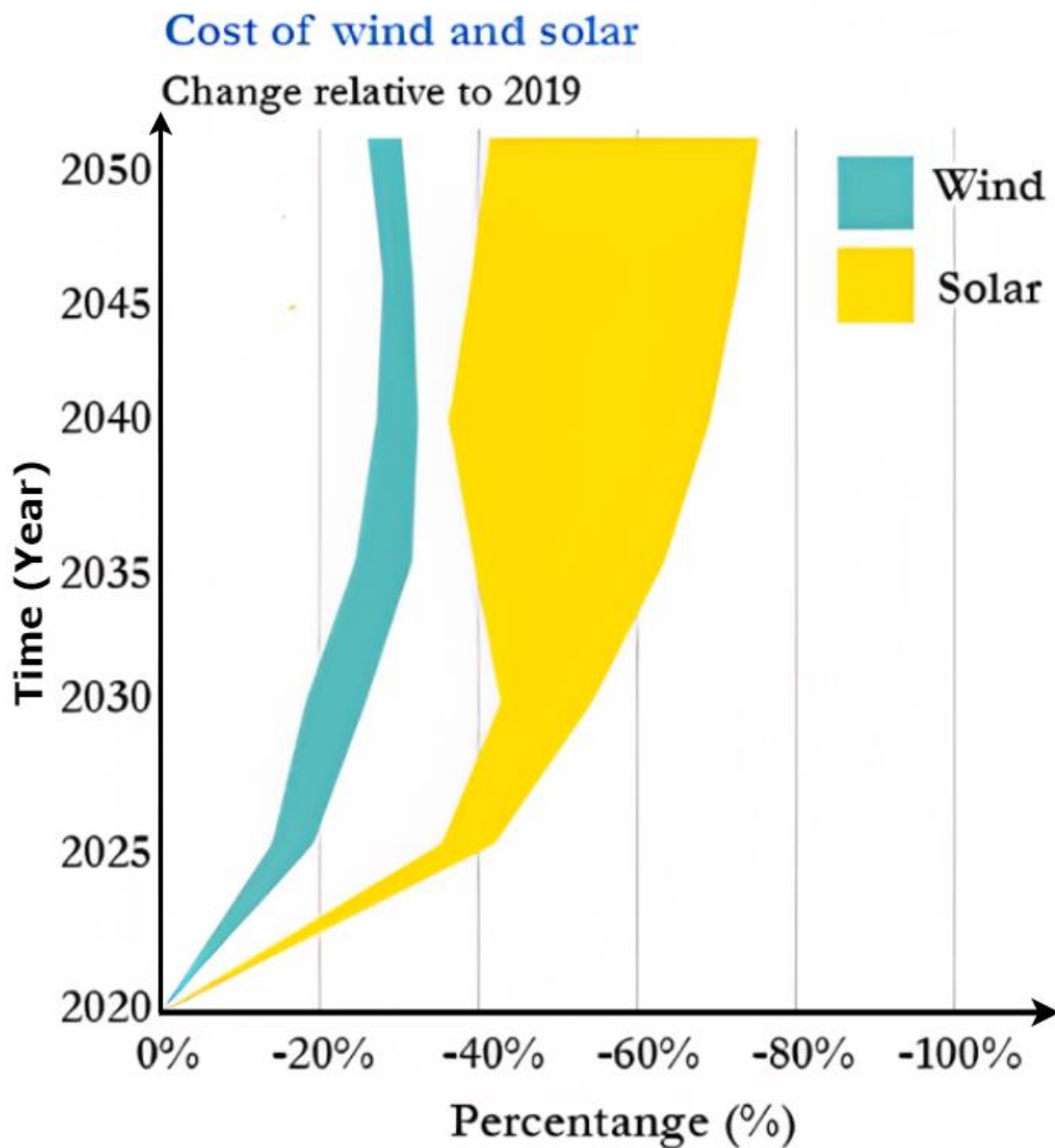


Figure 1. 3 Cost estimation of electricity generation from solar and wind energy in the future (Bp Energy Outlook 2022, n.d.)

In the Figure 1.3 shows us the estimated future installation and operating cost in the future, the cost calculation mentioned here is per Megawatt of electricity produced and as can be seen, solar panels have a great potential. (Bp Energy Outlook 2022, n.d.)

Solar energy has grown in popularity and economic viability. Over than 100 GW of photovoltaic panels were put into operation around the world in 2017. Simultaneously,

maintaining high energy generating efficiency, prevention and detection became an increasingly pressing topic. According to (Lillo-Sánchez et al., 2021) roughly 2% of photovoltaic panels are expected to break down after 11-12 years. Simultaneously, losses caused by dust collection (soiling) might be orders of magnitude more than losses caused by cell deterioration (Mejia et al., 2014). As a result, the significance of selecting the needed cleaning step cannot be overstated. Shade from growing plants can also cause losses. These and all similar problems can be prevented to a certain extent or damage is minimized and system efficiency is maintained by early diagnosis and regular maintenance.

Although the trend in the last decade has been a major shift towards the use of renewable energy, the most consumed energy sources worldwide are still dominated by fossil fuels. Driven by economies of scale and the opportunity for greater returns, the global solar industry is prioritizing the development of ever-larger, utility-scale projects. However, despite this trend, many solar project owners have given little attention to the advanced technology and data management software needed to effectively manage such large projects. Digital Twin technology, especially for large utility field solar, should be employed to reduce operating costs and improve asset performance. Although the solar operations and maintenance (O&M) business is already substantially digitalized, new digital technologies become practical as their costs fall.(Asimov, 2018) Monitoring systems currently collect and analyze massive amounts of data from large-scale solar plants, including energy meter data, irradiance, module temperatures, meteorological data, strings appropriate methodologies, and AC and DC side inverter parameters. As technologies advance, only data that has value and can be monetized will be captured, lowering data collecting costs and improving process efficiency. Advanced technology is also employed in the administration of operational information, such as actions, react and problem management, event kinds, ultimate cause diagnosis, and intervention cost (both in terms of labor and materials). The data is utilized to optimize the asset's operation at the lowest possible cost. (Y. Zheng, Yang, & Cheng, 2019) If digitalization, which is the integration of renewable energies into daily life, is carried out within a detailed and theoretically studied and created plan and integrates all these on a single platform, it will lead to keeping everything under control and becoming widespread. For this, it is controlled in a transparent, ethical way and it is important to establish predetermined rules. On top of all this, it should be considered that it can help the facilities that will use this system to connect with many things in real time or with very small delay times and offer the following advantages:

- Major industrial connection platforms may be used to improve asset lifecycle management and boost efficiency by leveraging big data. Systems may be modeled and simulated using digital architecture. Data may be analyzed at the portfolio, plant, inverter, string, and module levels. Benchmarking these corporate cloud solutions might be beneficial in the future.

- One of the main goals of digitalization tools and platforms is to increase the predictability and facilitate informed decision-making by leveraging the automation processes of renewable energy plants. In addition, isolated energy islands are avoided due to the interconnections provided by these systems, because these platforms are based on non-centered production.

- Proactive maintenance can utilize thermal and output trends to soiling, anticipate module degradation, system failures or component failures (e.g., module, inverter, or combiner boxes). This might avoid certain maintenance visits, predict others, and prevent costly unscheduled emergency visits. Modernization of sun panel facilities is necessary to make them more competitive and more efficient than traditional energy sources.(R. He, Chen, Dong, Sun, & Shen, 2019)

- Remotely sensed data and control, 5G and wireless connectivity, embedding test devices, including analysis of data can assist detect defects remotely, improve preventative maintenance, and improve plant production. Most solar data is being transferred via a standard interconnect DSI and IAn, with GPRS as a backup. Low-power wide-area networks (LPWAN) may be employed in the future, allowing for a wide range of IoT communications.(Solar Power Europe, 2017)

- Cloud computing is used to store data from data loggers on-site; however, an International Electrotechnical Commission (IEC) or comparable international standard is currently lacking in this area, and additional effort is thus required in this area, particularly given the potential savings.

- Digital field employees who have used mobile technology can improve ground operations, and command center units can track team positions to allocate them to assignments in the most efficient way possible. The more data field employees enter into systems in real-time, the higher

their monitoring on reaction times and the analysis of performance improvements. Smart eyewear technology might possibly be utilized to help with maintenance in the future.(Kufeoğlu, Liu, Karim, & Pollitt, 2019)

- Drones may be used for visual imaging of modules, wiring, and other aspect of project, as well as monitoring infrared thermal imaging. Drones of the future may feature on-board data analytics capability to spot trends and changes.

If look at what this can bring all this digitalization from a widely angle, we can briefly talk about the summing: All the data related to this product can be collected and analyzed in real time to ensure that any product is requested and its surroundings. It helps the modeling that will ensure that the air and market conditions required for the efficient production of renewable energy production to be predicted in the most precise accuracy. In addition, it will optimize the automation by optimizing the construction process with the analysis of other renewable resources and systems and the necessary information in the construction process, or by increasing the predictable production of production or increased production that can be obtained through these new resources. On the operational side, all these digital technologies will create new opportunities by facilitating the transmission and distribution of the power of the power that produces and creates a large development potential.(Madni, Madni, & Lucero, 2019) With automatic examination, it can be monitored by the vegetation of the land with automatic inspection - increasingly popularized agrovoltik concept for a vital importance for the agrovoltik concept-, as well as processing the thermal images to be obtained from the field with the help of AI, ensuring that a thing is taken to the maintenance before the deterioration, it will optimize the cost of operation and maintenance activities even more optimized. The convenience of remotely operating assets enables a degree of flexibility and management never seen before. Maintenance schedules are tailored to your equipment's actual life cycle. Less waste and limitless opportunities for industrial innovation: DT's provide improved performance, less maintenance, and the ability to focus attention on sustainability and improvement challenges that provide actual value to production.(Anderl, Haag, Schützer, & Zancul, 2021)

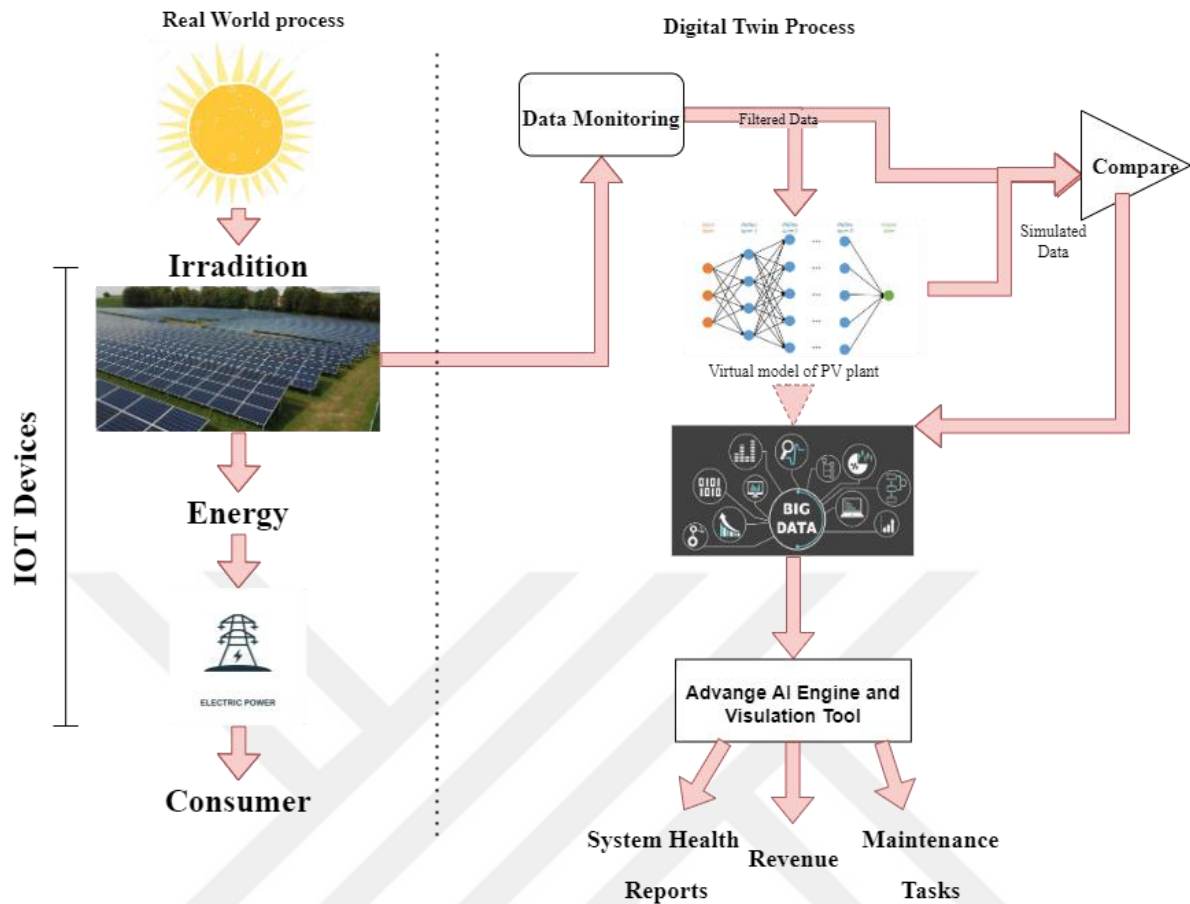


Figure 1. 4 Digital twin data flow and processing.

Installing a data gathering system that monitors various options of Photovoltaic (PV) tolls are a common technique of monitoring the health of solar plants the health of solar plants. This analysis is typically performed top to bottom instead of bottom-up, which would be too informative because more prone modules to error. The optimum tracking configuration collects data, controls each solar panel (Selmi & Belghouthi, 2017)(Vimalarani & Kamaraj, 2015), and makes this data available (Faranda & Leva, 2008) for analysis and prediction. However, the indirect character of the obtained information, along with the ever-increasing volume of time-based data (time series), provides a significant barrier for analytics. A typical data set representing solar panel functioning contains current, voltage, temperature, solar irradiation level, date and time of measurement. When all of these values are polled at five - minute intervals, the solar plants that will be operational in 2017 will yield 10 Gbyte per day. Weather fluctuation is another issue that affects the analysis. While solar panel performance is defined under standard test settings (STC), real-world applications exhibit considerable fluctuations in ambient conditions throughout the day. Solar irradiation, for example, is greatly influenced by meteorological pressure such as cloud, rain, snow, or mist. (Asimov, 2018)

Considering everything, there are two key issues that must be addressed in order to make business use of PV monitoring data:

- Collected data only indirectly represents system health (example: system supplies less power, but the reason is not clearly apparent); this demonstrates one of the shortcomings of top-down models.
- Amount of data continually grows over time; project analysis will be more complicated. The notion of PV Digital twin was proposed to alleviate the analysis's complexity. Digital Twin is a virtual replica of a real-world installation. It has the following major advantages: Examine monitoring data consistency; Use data mining to uncover and anticipate present and future problems.
- Make use of an AI Knowledge processor to help you make better business decisions. The virtual model that acquires knowledge and learns through data collected is the important element of the Digital twin. The model converts an almost infinite amount of time series data into a defined set of model variables.

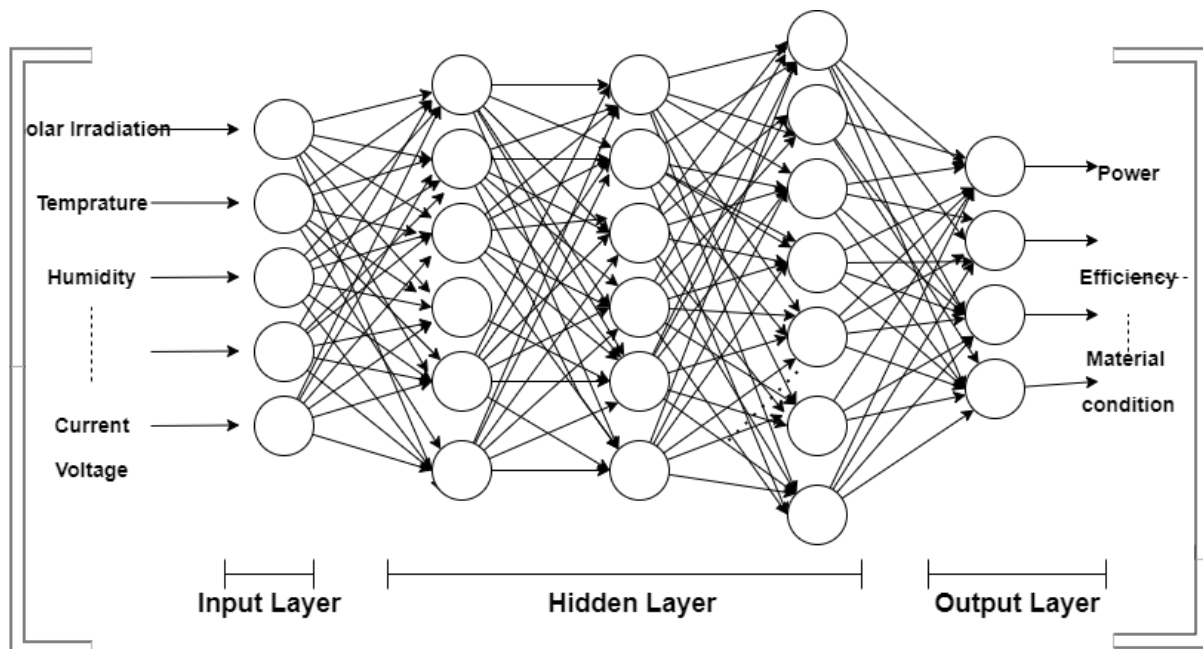


Figure 1. 5 Virtual representation of the solar panel model in proposed method

Based on the observed values of current (A), temperature (K), and solar radiation (W/m²), a theoretical model estimates the voltage (V) values at the endpoints of a certain photovoltaic panels. The learning procedure consists of modifying the model's inner factors to ensure optimum convergence of estimated and observed voltage levels. Once the tuning procedure is complete, model predictive variables (model parameters) may be utilized to calculate the efficiency of the PV.



2. LITERATURE REVIEW

Y. Zheng et al., (2019) The idea of DT and its attributes are interpreted in the realm of solar power planning in this work. A DT stages involved in the development for system lifecycle management in general has also been actively developed. In physical environment, all aspects of manufacturing and information sensing technologies are thoroughly covered. They designed three key function modules in the information processing layer: data storage, data mapping and data processing. This article tried to define the implementation process of completely parametric virtual modeling and the concept of constructing for DT application subsystems in virtual space. Finally, a DT scenario of a welded production line is built and investigated.

Asimov, (2018) In this study, the Digital Twin concept is focused on how it can be used in solar power plants. It is hoped that by developing digital twins of solar panels, integrating the Photovoltaic Mathematical model, Artificial Intelligence (AI) Information engine, and Big data analytics engine, the state of solar power plants may be monitored. This research specifically suggests using the opportunity provided by DT to identify answers to the sector's weaknesses, and the findings of the model developed for this purpose are discussed.

He et al. (2019) this article focuses on the greater accident rates and maintenance challenges encountered in the process industries as a result of complex architecture and demanding operating conditions. Security has an important place in systems implemented with a certain process; this study explains the data-based digital twin solution that combines all these in a single architecture to provide virtual modeling, process monitoring, pre-diagnosis and optimum control of the system. To represent closed-loop virtualization software and residual signaling with error-free state data when model parameters are unknown, an adaptive system concept is presented. The suggested digital twin system's efficiency and performance were assessed using various simulations on the Tennessee Eastman benchmarking process in the face of genuine failure situations. The suggested technique makes an essential contribution by establishing an integrated solution with digital twin principles based on algorithms developed to identify process faults and restore control parameters. The suggested digital twins enable controllers to make choices and optimize operations based on error data in the processing facility autonomously.

Fuller et al. (2020) The purpose of this article was to discuss the current state of Digital Twins, with IoT/IIoT and statistical analytics recognized as enabling technologies. The problems, uses, and innovative solutions for AI, Digital Twins and the Internet of Things (IoT) are discussed. A review of published resources on Digital Twins is made, resulting in a categorical review of available studies. The study categorizes them by research topics: manufacturing, health, and smart cities, and discusses these areas and various publications reflecting the current level of research. The preceding assessment identifies two further areas of significant interest: digital twins for medicine and smart buildings. As a result, the research adds to a comprehensive evaluation that encompasses not just manufacturing but also healthcare and intelligent buildings. This report covered each subject, emphasizing how academics are producing Digital Twins while also addressing problems and essential supporting technologies, so facilitating future work. It evaluated the technological solutions, difficulties, and public studies for Digital Twins.

Kufeoğlu et al. (2019) This article examines the current digitalization developments brought about by the advent of dispersed production; implementations of artificial intelligence, machine learning, deep learning, and Blockchain in the energy industry. Because of these tendencies, entrepreneurs and start-ups might offer new business models for existing big energy sectors. They attempted to assemble 40 business models from around the world that represented intriguing examples. While these organizations provide innovative and interesting products and services, it is suggested that many of them, like other start-ups, will fail in the long-term Sowing to flaws in its economic efficiency and revenue generating models. The energy sector faces new economic possibilities as well as new problems as a result of autonomous supply and an expanding number of producers. The advancement of digital technology has the potential to have a big influence on the energy business. Neural network, computer vision, and deep learning are frequently utilized in electricity trading initiatives to encourage digitization in electricity market, and its frameworks are frequently referenced.

Baboli, Babazadeh, & Kumara Bowatte,(2020) The goal of this study is to investigate the feasibility of integrating measurement-based control scheme with existing numerical modeling environment simulation abilities. The analysis in previous that such technologies have the potential to be integrated into real-time systems via Ad Hoc interfaces. It is common knowledge

that modeling the behaviors of physical objects is the first stage towards developing an entity's Digital Twin. Implementing models employing standard approaches (unique modeling of a physical link behavior for all elements) to match the functional needs of a Digital Twin is impracticable. The modeling technique for implementing the Distribution Network's Digital Twin must be less difficult than for individual modeling, but the generated models must not compromise the capacity to imitate physical entity behavior with appropriate performance. Second, it is noted that models should be capable of appropriate performance evaluation and adaptation. A dynamic identification approach that relies on Artificial Neural Networks (ANN) been offered as a solution, and it is recommended that it be utilized to identify the dynamics of linked models. However, the paper states that only a single model would not fulfill all needs, even the same criterion, under different situations. Combining ANNs with various nodes trained with different settings is said to get the best possible results for evaluating measurement results.

Pimenta, Pacheco, Branco, Teixeira, & Magalhaes, (2020) The present article has two main objectives: the first is to build a virtual twin of an onshore wind farm that will provide the Wind Turbine Operator with a tool for continuous observing of cumulative fatigue failure and consideration of alternative organizational processes; the second is to establish a consistent mathematical expressions of an floating wind farm to replicate experimental data to validate monitoring systems on such wind generators, to perform the preliminary tasks. U. S. National Renewable Energy Laboratory developed quick software to generate a simulation analysis of the onshore wind turbine (NREL). The building's structural and mechanical parameters, as well as the geometric characteristics of the tower and blades, were determined. After completely characterizing the wind turbine, the control methods were adjusted using data from supplier catalogues. Wind excitations for various operating circumstances are created, and structural reactions are estimated. It also gives a comprehensive characterization of the internal forces and calculation of modal characteristics at both sites, working in tandem with a sequence of extension and accelerating events at crucial spots along a real wind farm tower and blades under various operating situations. The goal was to detect and regulate the simulated and measured response, structural dynamic characteristics, and stationary and non - stationary internal stresses.

Nguyen, Tran, Besanger, Jung, & Nguyen,(2022) The digital twin integration approach is advocated in this article for evaluating decentralized renewable sources of energy.

Synchronization of the real-time simulator and the monitoring, control, and data gathering system results in the creation of the digital model of the electricity network. The digital twin could be incorporated into the bigger system by rearranging the topologies or by introducing simulated components without rearranging the topology. The installation is then finished with a strengthened cycle interface, which allows the physical integration of the genuine distributed energy source. Because the influence is then disseminated across the targeted distribution environment, it is examined both locally and internationally in this study.

Ebrahimi, (2019) The need of developing a digital twin concept of green energy generators is highlighted in this article. Extending the most recent technological analysis to the real simulation environment is unavoidable for 100% dependable green energy generation. A complete modeling approach that incorporates unpredictability and intelligent algorithm instruments to adjust the different scale modeling and platform is supplied to achieve this purpose. A network line was designed for this purpose based on the conceptual framework, beginning with the core and progressing in other dimensions. The concept is divided into three parts: multi-area physical models with expected values, multi-level smart techniques for model optimization, and unpredictability.

Salcedo-Sanz, Cornejo-Bueno, Prieto, Paredes, & García-Herrera, (2018) This paper focuses on selection issues that arise in renewable energy applications. The methods used to predict the system's generation may be extremely different, but they all have one common factor: they all accept input adjustments to enable the company to represent the system correctly. This is known as feature selection, and it is a major issue. This paper has two main goals: To begin, the most essential estimating methods for renewable energy applications involving property choice are being reviewed. In estimating systems, the study and debate of various property selection are considered. Wrapping FSP techniques demonstrate that they would be widely used due to their superior quality in this investigation. These structures contain a number of techniques that lead groundbreaking techniques. This potential was studied in the second half of the paper by giving a novel technique to characteristic recommendation based upon that Coral Reefs optimization technique, a new meta-nevristic. They believe that using this method, they may merge diverse search processes into a single method while still giving a high-quality worldwide selection mechanism. For the system's performance, actual statistics from wind turbines in Spain was utilized and compared with different regression techniques, as well as the wind rate prediction

issue from of the current numerical data. Advances of up to 20% in daily or hourly wind rate forecasts have indeed been observed using the suggested approach over techniques built without incorporating the characteristic selection procedure.

Lai, Chang, Chen, & Pai, (2020) The goals of this study may be divided into four categories. To begin, this article will aim to give an overview and critique of machine learning techniques used in renewable power prediction. Second, this research shows how to employ machine-learning methods for renewable energy projections, covering dataset pretreatment approaches, attribute selection methods, and forecast performance measurements. Third, renewable energy are examined, as well as mean percentage of error values and coefficients of correlation values. Finally, near the end of this paper, several prospective future employment prospects are explored.

Rahman et al., (2021) This paper examines machine learning (ML)-based techniques, particularly artificial neural network (ANN), in time series analysis forecasting challenges. To offer alternative resources to the globe, an "integration of renewable system" (IRES) comprising several solar and wind power could be deployed. The inherent complexity of green energy supplies, along with the neural network approach's smart capacity to absorb complicated time - series data inputs, prompted the adoption of ANN approaches in renewable electricity assessment. As a result, the purpose of this research is to investigate various data-driven ANN systems that can offer reliable predictions of renewable energy generation like solar, turbine, or hydropower.

X. Xu, Wei, Ji, Wang, & Gao, (2019) In this article, they analyzed four criteria to forecast the future growth of renewable energy in diverse regions: financial considerations, political issues, social aspects, and technological factors. Encouraging the production and the use of sustainable power is a latest fad in energy strategy in several locations. First, the globe has just been divided into seven area based just on Engineering News Record (ENR) geographic categorization (Asia-Pacific, Arab World, Canada, The Us, South America, Europe, and Africa), as well as the nation of sustainable power has been evaluated thus far: democratic, specialized, financial, and socioeconomic viewpoints. Second, the integrating prediction model is presented, which includes the differential autoregressive integrated moving average model (ARIMA), neural network model (NNM), support vector machine model (SVM), and one-way regression to

anticipate the advancement of solar and wind power sources. The outcomes are as follows: The predicted value using the differential autoregressive integrated moving average model (ARIMA) has been found to be somewhat optimistic; typically, the estimation would be at the maximum bound of the real value. The trajectory of the value derived from the neural network model (ANN) assessment is rather flat, typically maintaining near the lower limit of the predicted value. The estimation data generated by the support vector machine model (SVM) showed significant volatility. The Integrated model presented in this study combines the features of the three models and is thus used to obtain predictions that are more accurate.

J. Zheng et al.,(2020) The aim of this analysis is to assist to photovoltaic plant energy estimation by using the long-short-term memory and particle swarm optimization technique, as well as time series. Solar output forecasting is critical because it may be used as a decision-making tool in operation of the power system. Unlike previous research, this one aims at developing a model for predicting solar power production while photovoltaic systems are placed in more than one location. Particle swarm optimization is recommended to adjust the parameters of the long-short-term memory model in order to increase prediction accuracy. Furthermore, assessment was used to select the ultimate forecasting models by testing alternative long-short-term memory architectures. In the four scenarios involving distinct seasons, the standard error and r - squared errors of the suggested technique are the least among the estimating methods. The findings show that the suggested prediction model beats basic long-term memory, neural networks, and high gradients boosting in terms of prediction accuracy.

Shuvho, Chowdhury, Ahmed, & Kashem, (2019) This research looks at the effectiveness of an 80 KWp grid-connected solar farm in Dhaka. This research presents the development and application of a radiation from the sun forecast model fuzzy logic - based and artificial neural networks, with the goal of achieving high precision in a variety of meteorological circumstances. The precision of forecasted solar irradiance will influence the power production forecast of grid-connected solar PV that is critical for power distribution management and planning. NASA provided the data. When RMSE, R², and percentage precision are compared, ANN outperforms the fuzzy logic model. This study discovered that the ANN model outperforms the fuzzy logic method for predicting solar irradiance.

Motlagh, Mohammadrezaei, Hunt, & Zakeri, (2020) This paper examines the current literature on the use of IoT in power networks, and smart grids in specifically. Because renewable energy integration and energy consumption optimization are critical components of renewable energy transformations and mitigating climate change. Advanced techniques, like the Internet of Things (IoT), have a wide range of applications in the energy industry, including energy production, generation & distribution, and consumption. He highlights how IoT may be utilized to improve energy efficiency, promote the usage of renewable energy, and lower the ecological effect of energy usage. Also included are IoT supporting technology, such as numerous channels for cloud data processing and data analytics.

Badal, Das, Sarker, & Das, (2019) This article investigates the usage of alternative energy sources and power generating systems. Alternative sources widen the definition of the electricity system significantly. A microgrid's primary power source is renewable energy. This essay discusses the significance of microgrids in today's and upcoming worlds. A micro - grid system of energy systems that are linked in parallel by distributed generating units. The primary subjects of this article, which are addressed in depth, are the incorporation challenges of renewable energy and the management of the microgrid. The decentralized generating system's changeable nature causes frequency and voltage variation. Because the demand is uncertain, the behaviors are unmodeled. That unmodeled behavior has demonstrable implications on the microgrid's functionality. This paper studied the effectiveness of microgrid versus various situations and shares its outcomes. The power of the microgrid were regulated using several actuators, and the outcomes were investigated.

Martín et al., (2010) This paper compares predictive methods depending on the time series that are used to predict half-day amounts of world sun radiation across a 3-day time scale. The cumulative hour solar irradiation from morning to midday and midday to morning for each day is used to calculate half-day values. This ground cosmic ray's information utilized comes through AEMet sites in Spain. Autoregressive, neural network, and fuzzy logic models were evaluated. Because the half-day radiation from the sun data series is not fixed, this was determined that it was required to transform that into 2 additional fixed effects (sharpness factor and absent part), that were utilized as input data to the prediction models and were thus generated. Overall advancements in RMSD of the attempted models are discussed to the perseverance approach. The evaluation method demonstrates that almost all designs evaluated

enhance durability. Besides the Lerida station, in which modeling is based upon that wide angle ranking had little confusion because this amplitude does have a direct proportionality and is simpler to replicate to modeling techniques, neural network designs with failure elements as feedback are the finest tool for quantifying half-day values of cosmic rays.

Testad, Corbett, & Aarsland, (2013) This research provides a composite service model described as a multi networks for simulating systems having complex interplay among various elements. The height of towers, the scale of wind generators, and the proximity from coast, meteorological parameters, and other factors need to be addressed for investors to select the best maintenance approach. It is critical to investigate innovative maintenance approaches in order to lower the expenses and length of such operations. A multi-criteria decision system was created in this work to enable for the examination and selection of various maintenance techniques. An accounting standard that includes the cost of maintenance tasks, power losses, and real - time monitoring setup is provided. A simulator was created using NetLogo program for the objectives of this study, and empirical values are gathered. According to the findings, employing the composite repair method recommended in this paper may increase the performance of wind turbines while decreasing maintenance costs.

Gandoman et al., (2018) According to the outcomes, employing the mixed servicing method recommended in this paper can enhance the effectiveness of the wind turbines while decreasing operational costs. The objective of this essay is to offer a detailed review of FACTS/D-FACTS innovation, which is used in new electrical systems that employ RES alongside power electronic devices. Present Switches are experiencing significant modifications in designs based on intelligent grid-based quick existing approaches and computational intelligence control methods that make use of multipurpose voltage regulation, stabilizing, power factor correction, and interline power flow management. This publication included a range of FACTS-based approaches and controller design for dealing with non-linear/sensitive demands, alternative energies, power storage, and Vehicle chargers. The applicability and sufficiency of Facts controllers like SVC, D-STATCOM, SSSC, and APFs, as well as LC switching passive components, are discussed. The current analysis focuses on energy effectiveness, reducing losses, voltage stability, load conditions, power quality, and harmonics decrease at the identical connection with non-linear loads. Additional power electronics challenges are addressed, such as how to interface wind/PV fields using capacitive and inductive power network structures.

Hannan, Tan, Al-Shetwi, Jern, & Begum, (2020) This article will present a thorough examination of optimum controller techniques for the integration of renewable energy sources to microgrids, as well as various categorization in terms of design, attributes, performance, limitations, functions, expense, advantages and disadvantages. This research looks at new and standard optimization methods that are employed in microgrid applications. A thorough examination reveals that existing control optimization methodologies can enhance the management and utilization of renewable energy sources into microgrids. However, additional advancements in control management are required in the future to provide sustainable power management and alternative energy integration. Because most existing approaches are confined to theoretical calculations, experimental results is essential. The standard criteria for integrating renewable resources within microgrids, as well as the encouraging and harmonization, are emphasized. This report concentrated on many concerns, obstacles, and restrictions linked to the incorporation and management optimization of alternative energy sources, but also future studies ideas.

Hernández-Callejo, Gallardo-Saavedra, & Alonso-Gómez,(2019) In this paper some discussion of solar design process, management, and repair is provided. This had already been examined how the largest advancements in renewable technologies had concentrated on enhanced photovoltaic power architectures, and also maximum construction and operation that are crucial elements in solar photovoltaic study. Essential features and process simulation are examined in the context of PV system architecture. According to the papers researched on converters, hence the need for novel conversion designs. Those converters ought to be more capable of converting electricity. The goal is to provide fundamental standards for the development, construction, and servicing of solar PV based on information acquired out from functioning of 13 photovoltaic systems with varying features. All of the papers inside this research aim to increase productivity while lowering repair and operation expenses. It is claimed that this problem may be solved by changing the various aspects of a solar PV. Furthermore, authors mention the most typical difficulties encountered: inverters, regulation of the quantity of the power generated the issue of loss in performance of units because of climatic conditions (dust & shade), the issue of hot zones and snails track creation.

3. MATERIALS AND METHODS

In this study, two different programs were used, the first of which was Matlab, where the Simulink and modelling part was made, and the second one, python, where the Machine learning process part was carried out. It is the most preferred program especially in modelling with its Matlab product and working system. Although there are many different programming languages in machine learning, python programming language stands out with its simple and wide library. Another reason why the Python programming language is preferred is that it facilitates working with big data. In the following sections, these programs will be explained in detail and information will be given about how they are used in our study.

The circuit diagram created by the selected model is compatible with the equivalent schemas of the parts, and the reason for its preference is that the number of unknown parameters increases with distance from the ideal form. However, the datasheets of most manufacturers do not provide enough information about the factors affected by weather conditions (irradiation and temperature). As a result, in addition to the use of the information provided by builders, some assumptions about the physical nature of these parts are required to develop the Matlab/Simulation model of the PV module and all other electronic parts. The main purpose here is to perform and consider all the processes from the beginning to the stage of electricity generated with the help of solar panels as a single process as in real life. Therefore, the aim is to try to keep all the different processes that take place in the system at the same time in the most precise, fast, and therefore the closest possible result in the field, while keeping it in the most efficient state possible at any time.

3.1 Matlab/Simulink

A grid-connected PV system typically includes of photovoltaic panels, a DC/DC converter, an MPPT controller, an inverter, and power grid equipment. This has no power storage loss since no batteries are needed since it is not a stand-alone system. Matlab/Simulink software is used to model the system's components. Matlab/Simulink is chosen for such systems because of its scalability, reusability and adaptability.(Zainal, Ajisman, & Yusoff, 2016)

It is divided into 3 main parts to keep the solar panel system under control more easily. The first of these is the solar panel section where electricity is produced by solar panels, the converter section, which is the second stage where the direct current produced in this section is converted to higher voltage with a DC/DC converter, and lastly, the voltage is increased once again with the help of transformers in order to switch from direct current to alternating current and give it to the grid. There is the 3rd part where it is upgraded. However, you should not forget that all these stages are directly related to each other and occur at the same time under normal conditions, and this is exactly why this work runs the simulation in one piece and then evaluate the results as if they were three separate pieces.

As mentioned before, the main reason for the 3-part analysis is to make it easier to determine if there is an error in any part of the system or if there is a loss of efficiency for some reason, and to help the system analysis by making it easier to understand what kind of reason it may have been.

In European countries, solar panel systems should be separated from the grid just in case of some error in the system or overcharge. In this design DC/DC converter is used as a bridge also it gives the flexibility to connect the system to Uninterruptible Power Supply (UPS) if it will be needed. In this study, main idea is to create a perfect program that can work with every type of solar power plant, so the design of the plant was chosen because of this. This plant started production at the beginning of this year in Sevilla within the scope of this project. On the internet you can find different version of this design but this one created as part of this project.

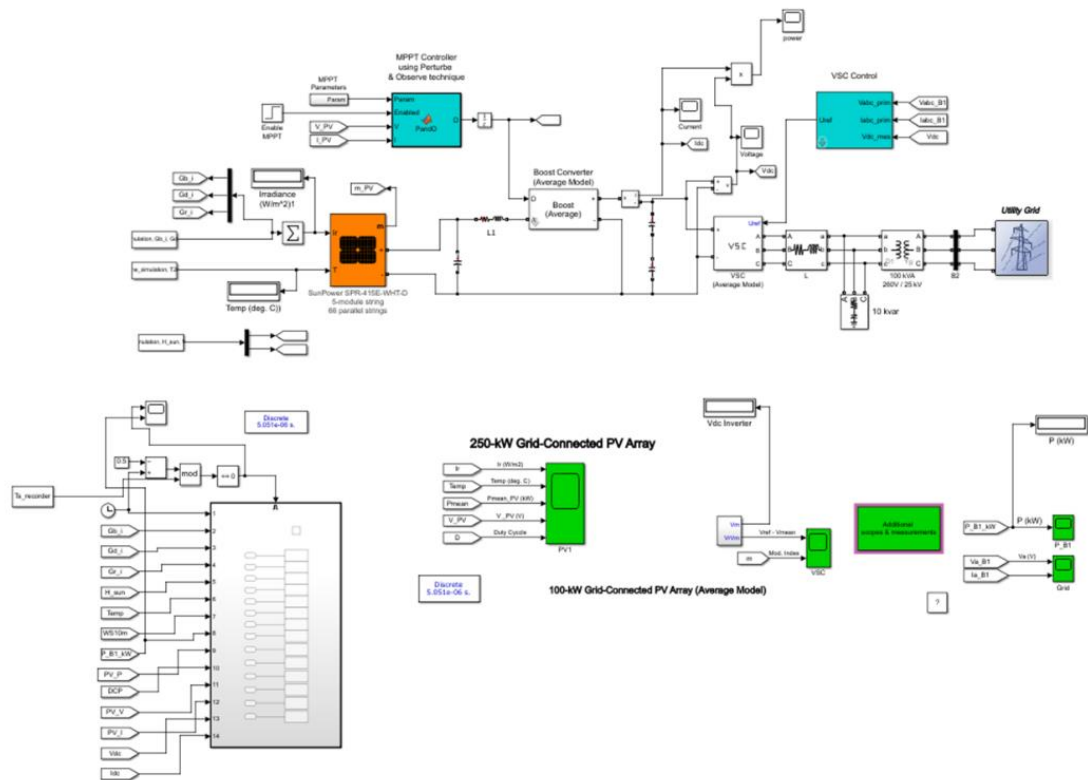


Figure 3. 1 Simulation Model of the Solar Panel System used.

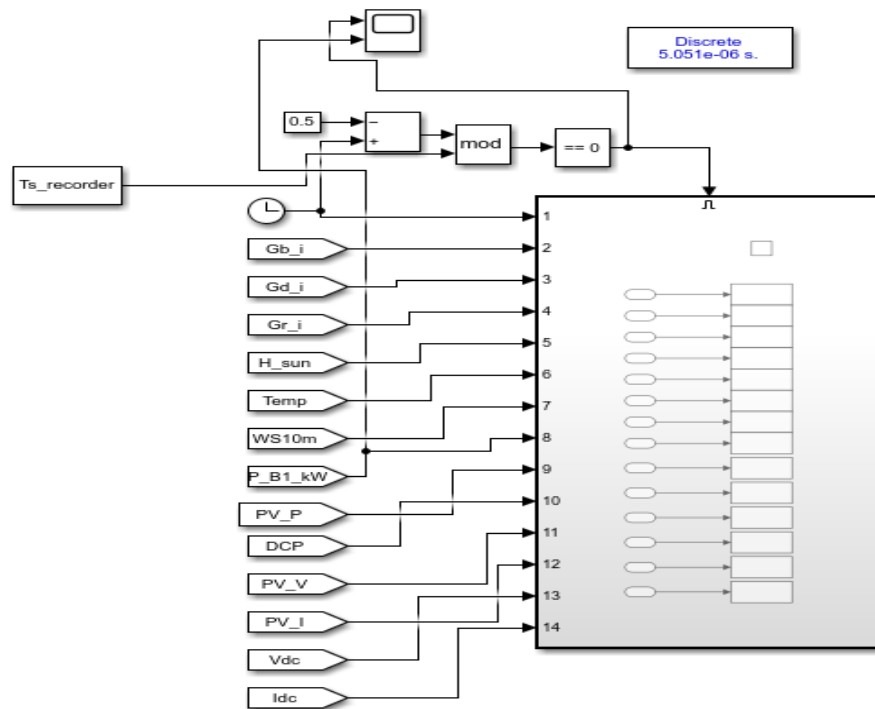


Figure 3. 2 The results obtained in the simulation can be transferred as a table thanks to this diagram.

The equipment used in the simulation model is generally used with reference values so that it can give the closest result to reality. The values you see in the input section are accurate measurements made over the course of a year of monitoring a specific area, and they were sourced from the European Union Commission (JRC Photovoltaic Geographical Information System (PVGIS) - European Commission, 2016).

Table 3. 1 The data we obtained as a result of the simulation.

Gb(i)	➤ Amount of beam (direct) radiation occurring in the inclined plane (array plane) (W/m2)
Gd(i)	➤ Amount of scattered radiation (W/m2) in the inclined plane (array plane)
Gr(i)	➤ The amount of radiation reflected in the inclined plane (array plane) (W/m2)
H_sun	➤ Specifies the height of the sun (degrees)
Temp	➤ Air temperature at 2m altitude (degrees Celsius)
WS10m	➤ Total wind speed at 10 m height (m/s)
P_B1_kW	➤ The power produced in the system before it is given to the grid (W)
DCP	➤ Power value obtained after DC-DC Converter
PV_P	➤ Power value obtained from PV panels
PV_V	➤ Raw Voltage value obtained from PV panels
PV_I	➤ Raw Current value obtained from PV panels
Idc	➤ Raw Current value obtained from DC-DC Converter
Vdc	➤ Raw Voltage value obtained from DC-DC Converter

3.1.1 Solar cells

Solar cells are a non-linear device and consist of materials that absorb light photons and release electrons with the energy they obtain from them. The absorption of light directly and converting it into electricity is called Photovoltaic event. Solar cells can be represented as a current source in circuit, which you can see in Figure 3.3. The current source of the solar cell can be expressed

as the dependent source, which consists of a single diode, a shunt resistor R_s and a series resistor representing the p-n junction cell.(Bellia, Youcef, & Fatima, 2014)(Selmi & Belghouthi, 2017)(Vimalarani & Kamaraj, 2015) R_{sh} describing an internal resistance of cell to the current flow. The terminal current of a solar cell may be estimated as a function of photo-generated current, photodiode current, and shunt current as follows:

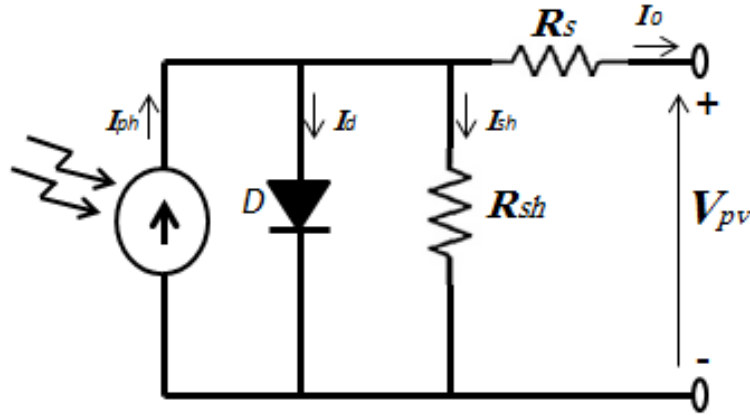


Figure 3. 3 Circuit diagram of solar panel (Bellia, Youcef, & Fatima, 2014)

$$I_o = I_{ph} - I_d - I_{sh} \quad (1)$$

Where:

I_{ph} represents the current produced by the incident light,

I_d represents the current through the diode.

I_{sh} represents the current through the parallel resistor.

R_{sh} is the relationship of current-voltage characteristic of the PV array is explained by the following equation:

$$I_o = N_p * I_{ph} - N_p * I_{rs} [e^{(q(V + R_s * I_o) / (A_k * T * N_s))} - 1] - N_p * q(V + R_s * I_o) / (N_s * R_{sh}) \quad (2)$$

where:

I_{rs} represents the reverse saturation current inside the photovoltaic cell.

V represents the output voltage (V) from the photovoltaic cell

A represents the diode ideality constant.

T represents the operating temperature of the reference photovoltaic cell,

q represents the electron charge ($1.60217646 \times 10^{-19}$ C),

k represents the Boltzmann constant ($1.3806503 \times 10^{-23}$ J/K),

R_s and R_p are the series and shunt resistances of the cell, respectively. The following equation relates to the photocurrent I_{ph} produced, ie solar irradiation:

$$I_{ph} = \frac{G}{1000} * (I_{sc} + k_i(T - T_r)) \quad (\text{Bellia et al., 2014}) \quad (3)$$

Sunlight transmitted in the form of energetic particles known as photons. The creation of electric charge occurs within the PN junction's depletion layer. Like early discussed with the diodes, this depletion zone is located around the PN junction which electron first from N-type silicon had migrated into the gaps of P-type semiconductor. Whenever a portion of the energy is captured by this kind of N-Type doped silicon, this displaces a particle, resulting in unpaired electrons as well as a gap. Both unbound electrons and holes have enough energies to escape in depletion region. The transfer of electrons takes place via a kind of wire connected from the cathode (N-type silicon) to the anode (P-type silicon) (Figure 1.6). A particle gets drawn to a positively charged ions of the P-type semiconductor then passes via the outside charges (meters), causing the electricity to pass. This dislocated electron's holes being drawn toward the negatively charged of the N-type semiconductor and relocates to the back electrochemical reaction. These electrons connect with holes when that penetrates the P-type quartz from back electrical conductor, reestablish electroneutrality.

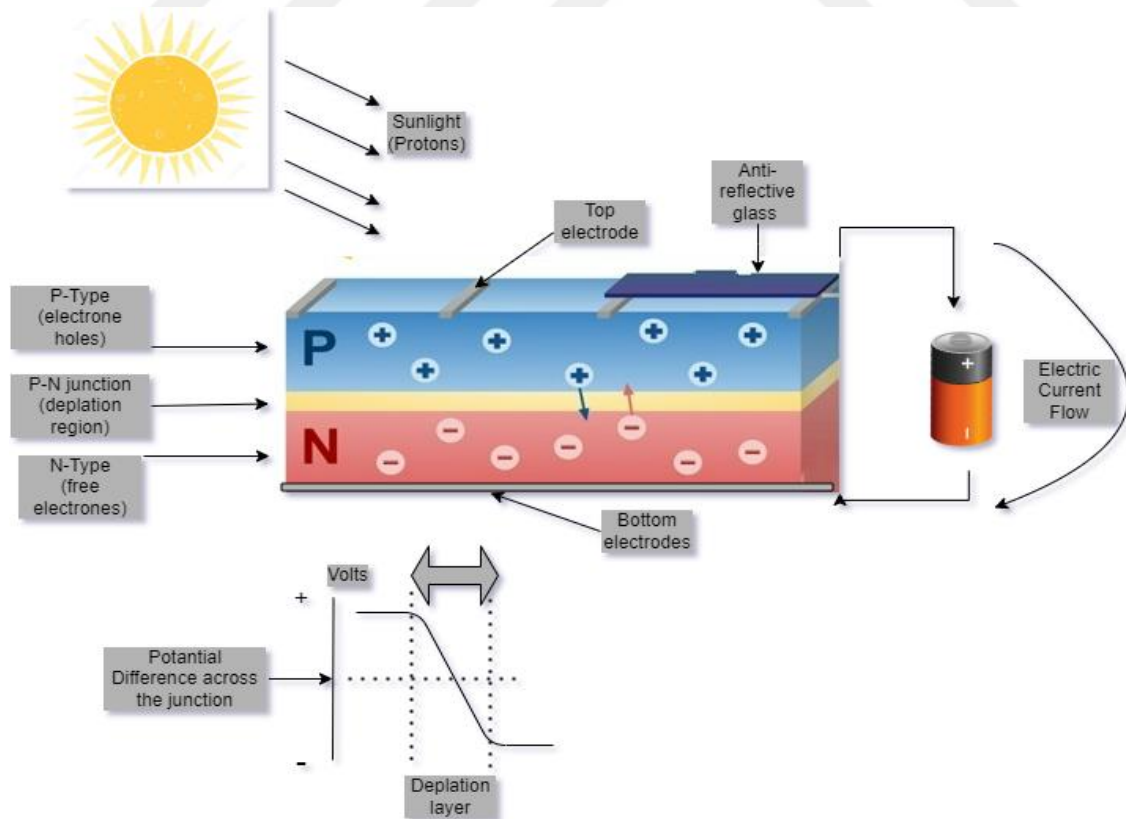


Figure 1. 6 Basic PV cell design

There are different approaches for solar cells, single-junction solar cell type is preferred in this research. Because it is not yet ready for buying commercial solar power plant because it is currently under evaluated and investigated. Differentiating among various types of solar sometimes entails categorizing them as first, second, or third generation panels. Every one of those solar cells is one-of-a-kind and has a distinct visual aspect. General information about every solar panel type will be provided here.

First generation solar panels are the classic varieties of solar cells composed from monocrystalline silicon or polycrystalline that are typically utilized in traditional settings.

3.1.1.1 Monocrystalline silicon solar cells

Monocrystalline silicon solar cells are the most produced and earliest form of solar cell. This monocrystalline solar panels are constructed from around 40 monocrystalline silicon solar cells. These solar cells are composed entirely of silicone. A silicone crystal is put in a tank of boiling silica during the process of production (known as the Czochralski technique). The crystalline can then be carefully drawn out of the tank, enabling the liquid silicone to form a robust crystal casing around it known as an ingot. After that, the ingot is precisely cut onto thin silicon boards. The board is converted into cells, and the units are then connected to make solar modules.

Due to the obvious method, lighting reacts with pure silicon wafers, monocrystalline solar cells are clearly spotted by their curved corners and uniform black appearance. Whereas the units are black, the backing sheets and casings come in a range of styles and colors. The monocrystalline cells have a square form with the edges eliminated, resulting in tiny gaps spaces between the units. Monocrystalline silicon photovoltaic modules are a high-quality panel. While monocrystalline panels seem to be more costly, it has a longer life expectancy and greater efficiency. Because the cells are made of single crystalline structure, it produces more energy. Additional benefit is that they are considerably lower impacted by hot temperatures than polycrystalline cells.

3.1.1.2 Polycrystalline silicon solar cells

Polycrystalline panels, sometimes known as "multi-crystalline" modules, are frequently regarded as the half -range panel. Polycrystalline solar cells are the less expensive choice but being less effective. Polycrystalline cells, as monocrystalline panels, are constructed of silicone. Polycrystalline cells, on the other hand, is created by multiple (Poly) pieces of silicon crystal fused collectively. As a result, polycrystalline solar panels are less efficient and have a shorter life expectancy. Polycrystalline solar panels are a recent discovery among the many varieties of photovoltaic modules, but they are growing in popularity and effectiveness swiftly. Polycrystalline cells, as monocrystalline panels, are constructed of silicon. Polycrystalline units, on the other hand, are created by molten silicon crystal pieces. The silicon crystals are put in a tank of boiling silicon during the production operation. Rather than carefully drawing it out, crystalline is permitted to shatter and cool. After cooling in its mold, the fractured silicon is thinly cut in polycrystalline solar wafers. These wafers are combined to produce polycrystalline modules. This results in a reduced actual cost however it also reduces efficiency (approximately 15%), less produce effective, as well as a shortened lifetime since they are more impacted by high heats.

Due to the way sunshine bounces off the crystalline, polycrystalline cells look blue. Light bounces significantly from silicone pieces than those from a clean semiconducting material. The back casings and casings are typically silver with polycrystalline; however, this might vary. The unit is squared, also there is no spaces between both of the unit's edges.

3.1.1.3 Thin-Film solar panels

Thin-film solar panels, as opposed to monocrystalline and polycrystalline solar cells, are thin, elastic, and shallow in appearance. It is due to the fact that the cells inside the modules are about 350 times thinner than the crystalline wafers being used monocrystalline and polycrystalline photovoltaic panels. However, thin-film frameworks can be quite big at points, giving the complete solar system the aspect of a monocrystalline or polycrystalline plant. Thin-film units could be azure or black due to the materials used. Thin-film solar cells are created by layering semi - conductive substances like silica, cadmium telluride, and copper indium gallium selenide. The semi-conductor surface is sandwiched between two transparent electrode sheets, with a glassy coating on the surface to assist catch sunshine. While silicone is utilized in the

fabrication of thin-film solar cells, this is not the same as hard silicon wafers. It is, instead, a non-crystalline silicon.

When compared to crystalline panels, thin-film solar panels have poorer efficiency and power capabilities. With efficiency of roughly 11%, they require a huge amount of roof space to generate a large amount of clean power. It is generally prone to decay faster than crystalline panels, leading in the lowest warranties.

3.1.1.4 Cadmium telluride solar cell

In the several forms of photovoltaic arrays, the photovoltaic approach employs Cadmium Telluride, which allows for the fabrication of solar cells at a low cost and hence a shorter payback period (less than a year). This is the solar energy technology that requires minimal quantity of water to produce. CdTe solar cells can reduce its environmental footprint as much as possible due to their fast power repayment period. The sole downside of employing Cadmium Telluride is that it is hazardous when consumed or breathed. This is amongst the most difficult obstacles to overcome, particularly in Europe, where several people are hesitant about employing the technology underlying this sort of photovoltaic panels.

When compared to crystalline panels, thin-film solar panels have poorer efficiency and power capabilities. With efficiency of roughly 11%, they require a huge amount of roof space to generate a large amount of clean power. It is generally prone to decay faster than crystalline panels, leading in the lowest warranties.

3.1.1.5 Concentrated pv cell

Concentrated photovoltaic cells create electricity in the same way that regular solar installations do. These multi-junction photovoltaic arrays have an effectiveness rate of up to 41 percent, which is the greatest between all solar photovoltaic installations so far. The title of such Concentrated PV modules refers to what causes these particularly effective in comparison with other methods of photovoltaic arrays: curve mirrored floors, optics, and perhaps even cooling equipment is utilized to compress the amount of solar radiation and therefore boost their performance.

Concentrated PV cells have therefore been one of the most successful photovoltaic cells, with such a remarkable design and performance rate as high to 41%. How this stand is this such Concentrated PV solar cells will only be as successful when it tracks the sunlight at exactly the right degree. A solar tracking system within the photovoltaic panels is accountable for implementing the sunlight in terms of achieving exceptionally high-performance percentages.

The digital twin concept can play a very important role in the installation and monitoring of especially 2nd and 3rd generation solar panels. Because these systems are more innovative and considering that some of them are still in the research phase, it is important to follow up and examine them in detail. In this way, the discovery of new techniques can be opened.

3.1.2 Conventional perturb and observe method

Due to the low energy efficiency and non-linear energy graphs of photovoltaic arrays compared to fossil fuels, monitoring a Solar Panel's maximum power point (MPP) is one of the most important steps in PV systems.(Faranda & Leva, 2008) The P&O technique has a slow but dynamic response system, as long as there is a slight increase in the sample. The biggest problem of P&O is the steady state error that causes continuous oscillation up and down around the MPPT when it reaches the MPP point.(Sera, Mathe, Kerekes, Spataru, & Teodorescu, 2013) A small increment is needed to reduce this error. This results in a regulation voltage that deteriorates with each MPPT time. The fluctuating motion that occurs around the MPP, fluctuations in photoelectric systems cause power loss. This phenomenon usually occurs in slowly or steadily changing atmospheres. In this study, the proposed ANN system was trained with the data obtained from P&O MPPT. (Thomas & Nithiyananthan, 2018)(Nicola Femia, Petrone, Spagnuolo, & Vitelli, 2009)(N. Femia, Petrone, Spagnuolo, & Vitelli, 2004) The working algorithm of P&O Systems is given in figure 3.4:

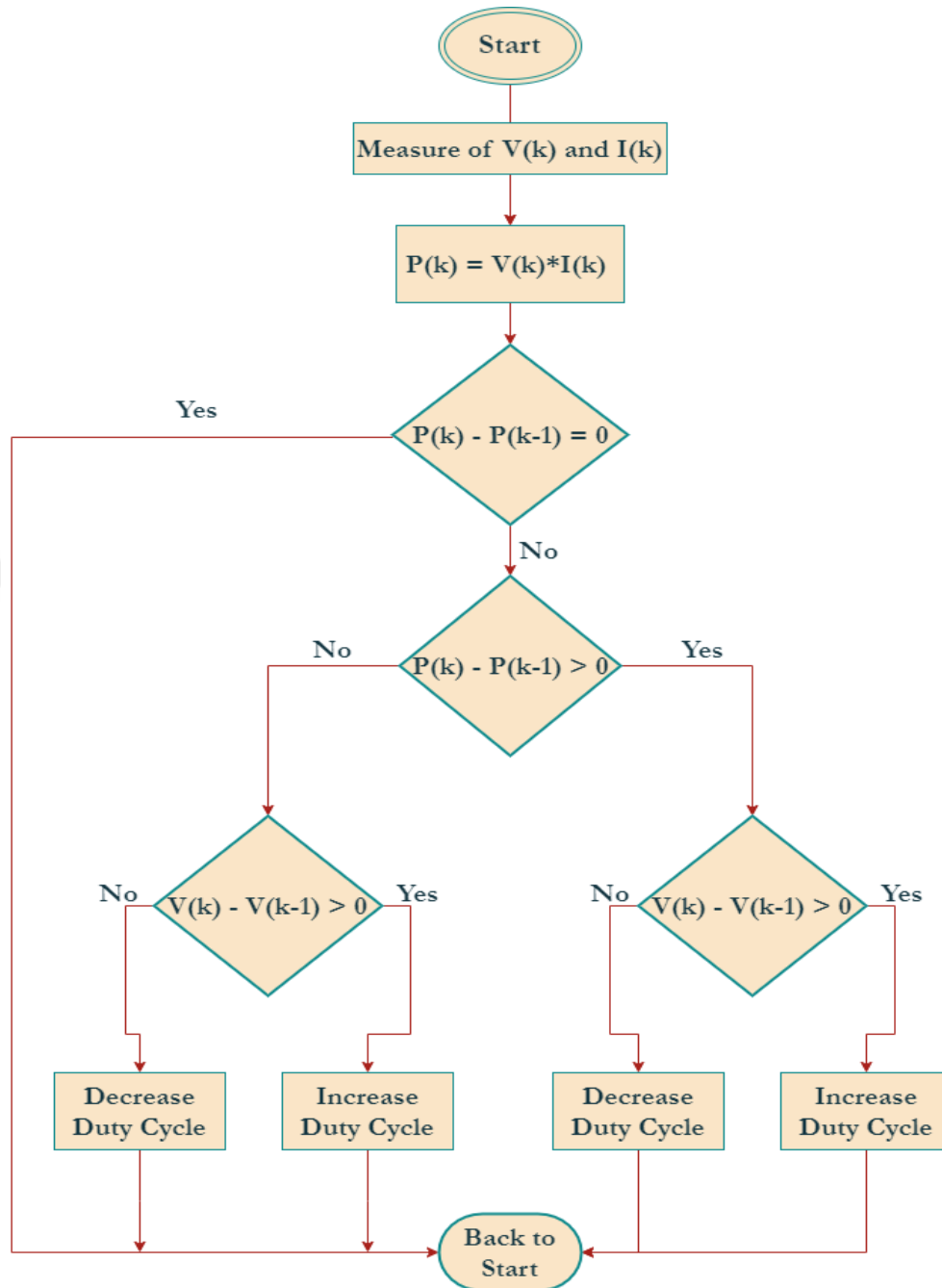


Figure 3. 4 Flowchart of the conventional P&O algorithm

3.1.3 DC/DC converters

In regulated switching power DC power supplies, DC/DC converters are commonly utilized. These converters' input is an uncontrolled DC voltage obtained by a PV array; hence, it will fluctuate owing to variations in solar irradiance and temperature. Although the input voltage changes, the average DC output voltage have always to be managed in these converters to be equivalent to the required value. In terms of energy, voltage output regulation in a DC/DC

converter is accomplished by continually regulating the quantity of energy taken from of the source and pumped into load, which is regulated by the relative lengths of the absorption and injection periods. A switching cycle is made up of these two fundamental operations of energy absorption and injection. Intuitively, if the converter's energy storage capacity is too little or the switching time is too lengthy, the converter will have sent all of the energy stored to the loads before another cycle will begin. This creates an idle time directly following the injection period, wherein the converter does nothing. As a result, the converter may function in two modes depending on its own energy storage systems and the duration of the switching interval. The discontinues conductivity and constant modes are the names given to these two modes.(Alagu, Ponnusamy, Pandarinathan, & Mohamed Ali, 2021)

It is always preferable in a DC/DC converter that the output voltage, V_o , remains constant regardless of changes in the source voltage, V_s , the load current, i_{Load} , and fluctuations in the element values of the converter circuit. These disturbances can be caused by, among other things, second harmonic regular intervals variations of an off-line power network produced by the bridge rectifier and tried to apply to the DC/DC converter, variations in the source voltage V_s caused by switching (on/off) of nearby power load demand, and variations inside the load current, i_{Load} . DC/DC converters are classified into six types: Boost, Buck-Boost, Buck, Cuk, Sepic, and Zeta. Amongst all converters, the Boost converter, a step-up converter that gives a greater voltage at the load side, V_o , than the source voltage V_s , is the most extensively used DC/DC converter. The Boost converter's open-loop way of operating has poor voltage control and an unfavorable dynamic response.(Del Moral et al., 2015)

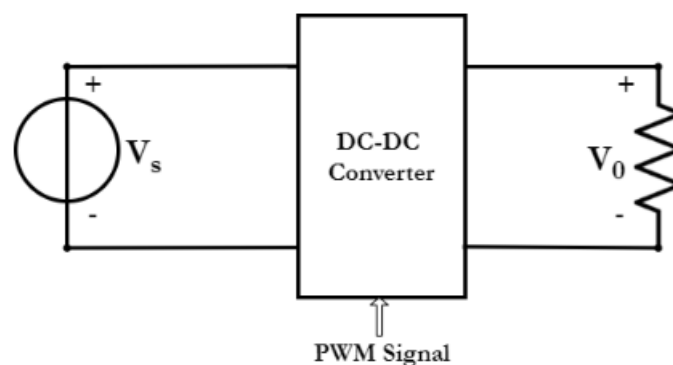


Figure 3. 5 Block Diagram of a DC-DC Converter

In this research, the boost converter, which is used because of its simple and efficient design, was used. In this research, the basic working principle of the boost converter will be explained in order to understand why this type of converter is chosen.

3.1.3.1 DC/DC boost converter

Using a DC/DC boost converter, a low input DC voltage is transformed to a higher DC Output voltage. Because the voltage source is ramped up in comparison towards the output voltage, this is also known as a step-up converter. In theory, semiconductor power switches and standalone electrical & mechanical elements could be used to create DC/DC converters.

The converter in a DC/DC converter function in two modes:

- Mode of Continuous Conduction
- Mode of Discontinuous Conduction

The diagram (Figure 3.6) depicts the DC/DC boost converter's continuous conduction mode circuit, which includes a coil, capacitance, switches device, diodes, and input voltage source. The pulse width modulator (PWM) is used to regulate the boost converter circuit switch. When this switch is turned on, electricity is generated inside the coil, and therefore additional energy is transferred to the terminal. (Ghabeli Sani, Mohammadi, Banaei, & Farhadi-Kangarlu, 2019)

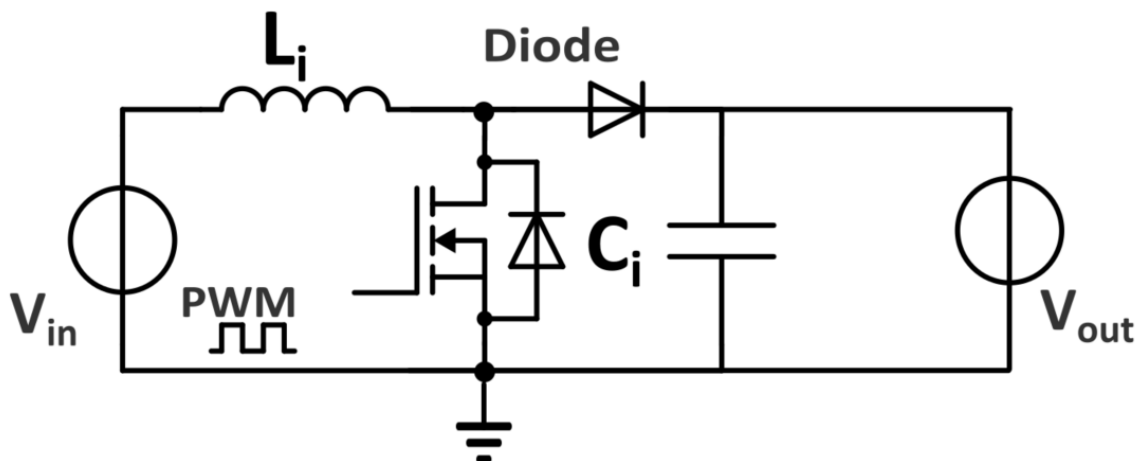


Figure 3. 6 DC/DC Boost Converter basic circuit diagram

The DC/DC boost converter's discontinuous conduction mode circuit includes parts such as a capacitance, inductance, reference voltage, diodes, and switching device. If the device is

switched on with this discontinuous conduction mode, power is supplied to the energy storage component, the coil. If the device is switched off for an extended length of time, the current in the inductor will drop to zero till another switching period. As a result, the capacitors charges and discharges in relation to the input power. However, in this case, the voltage output in discontinuous conduction mode is smaller than that in continuous conduction mode.(Mirzaee & Moghani, 2020)

3.1.4 Solar pv-grid system

Solar panel system, inverter, a power conditioning unit, along power grid components comprise a grid-connected PV system. It effectively utilizes solar energy generated electricity since there are no power storage inefficiencies. When the conditions are favorable, the grid-connected PV system sends extra electricity to the utility grid that is not consumed by the associated load. Nonetheless, in isolated systems, batteries are employed to store energy, or energy is directly linked to the load.(Abdelsalam, Massoud, Ahmed, & Enjeti, 2011)(Akinsipe, Moya, & Kaparaju, 2021)

The transformer (Figure 3.6) is installed to the PV system to link it to the nationwide high voltage grid. Its role is to modify the voltage-to-current ratio given from one voltage source to match the character traits of the load while maintaining power constant. Transformers are made up of three parts: the primary coil, which acquires energy from the AC source (in this case, the inverter), the secondary coil, that also gets power from the main winding and provides it to the load (in this case, the national grid), and the core, which provides a lines of magnetic flux. (Bhandari et al., 2014)(Karthikeyan, Rajasekar, Das, & Karuppanan, 2017)

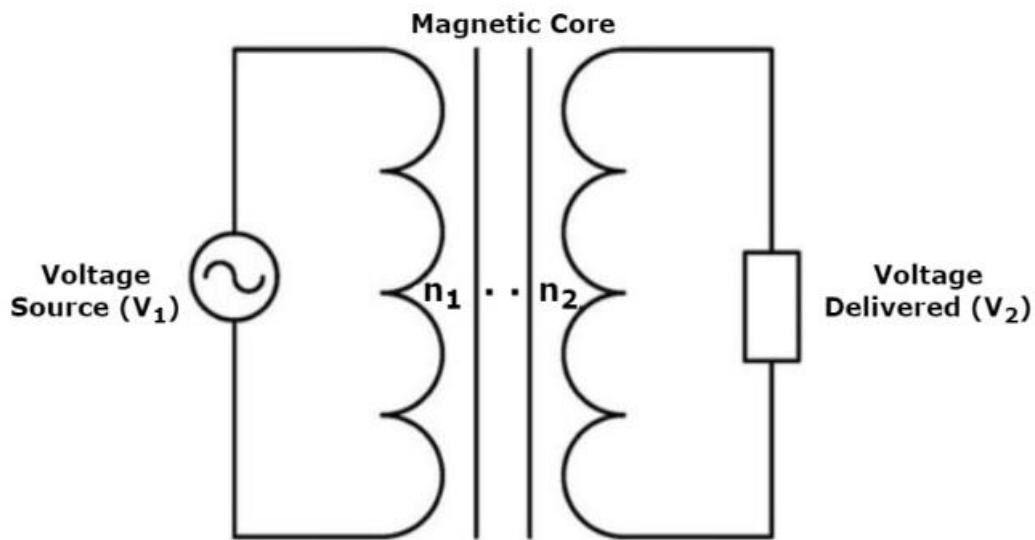


Figure 3. 7 Ideal Transformer Schema

A transformer's basic operating principle is the ratio of faraday's law of induction between primary and secondary windings connected by a magnetic flux. The basic transformers is made up of primary and secondary windings that are electricity separated yet magnetically coupled via a reluctance channel. To prevent or minimize eddy current loss in the transformer, the base is built of laminating plates. This laminating plate are linked using staggered joints to prevent the free air space running across the cross sections. These staggered seams are referred as imbricated. The mutual inductance of both coils is quite high. When one of these coils is coupled to an alternate supply voltage, a bilateral EMF is created in another coil. The Faraday's law of mutual induction helps explain it.(Leela Nivashini, Madhumitha, Sindhu, & Maheswari, 2020)

A magnetic field connects the electrical windings of a transformer. In special-purpose transformers, the coupling rises as the magnetic core force increases. Transformers are not constructed with the primary winding across one side and the secondary winding on either. Whenever an alternating current voltage is applied to the primary coil, magnetic flux is generated that causes voltage to be generated in the secondary coil. Once the flux is curved, the secondary coil produces a sinusoidal voltage. The main is linked to the generator, and the second is linked to a consumer. Facets with lower or higher voltage are offered. This magnetizing current is reduced by building the core using low reluctance substance.

Transformers are either step-down (giving lower voltage to the load) or step-up (giving higher voltage to the load) according to the ratio of the number of turns in the secondary part and the primary part to each other.

$$V_2 = (n_2/n_1) \times V_1 \quad (4)$$

$$I_2 = (n_1/n_2) \times I_1 \quad (5)$$

$$P = I_1 \times V_1 = I_2 \times V_2 \quad (6)$$

Where I_1 , V_1 are the current and voltage delivered from the source to the main coil; I_2 , V_2 are the current and voltage provided to the load by the secondary coil; and n_1 , n_2 are the primary and secondary winding turns. The anticipated high voltage grid comprises of a distribution network and a transmission network.

3.3 Machine Learning

It cannot directly say that a machine-learning model is best because the systems used in general are different from each other or because they are used for different purposes. Therefore, will try more than one model using the test data have, and aims to obtain the best working estimation algorithm with the system, considering the results it has obtained. As mentioned before, the reason for using regression is to predict the face that can be obtained based on the Input data. The difference between the predicted and actual if it is more than 10% wrong, it is intended to inform the user that the efficiency is lost. The main purpose to be done here is to observe our site instantly and keep it in the most stable and efficient way possible.

If were working on classification problem, there are two main indicators that should look at: accuracy and validation loss values. Nevertheless, when doing Regression, the situation is a little more complicated because each parameter draws attention to a different issue and directly affects the performance of the model. So even if the prediction results of the model are very good, if the error rates are high or inconsistent, it may mean overfitting or that the model is not good enough. For all these reasons, in this section it will look at the three main error values and the prediction results. Firstly, going to explain the errors and its meaning for machine learning and then the results will be shared.

Mean Absolute Error: Absolute error in machine learning refers to the extent of the difference between the forecast of an observation and the real value of such an observable. MAE calculates the amount of errors for a collection of predictions and measurements by calculate the mean of absolute errors. MAE is also known as the L1 loss function. MAE, being one of the most frequently utilized loss functions for regression issues, assists users in turning learning issues into optimization problems. It also provides as a simple quantitative error assessment for regression difficulties.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - x_i| \quad (7)$$

The symbol Σ represents summation, where y_i and x_i are the actual and calculated values, respectively, for the i th observation. N represents the total number of observations.

Mean Squared Error: MSE is the squared error average, which is used as a loss function in least squares regression. A model's mean squared error with regard to a test set is the average of the squared prediction errors over all instances inside the test set. The prediction error for an instance seems to be the ratio between the real and expected values. It is the total of the square of the ratio between anticipated and real target variables divided by the number of pieces of data over all data points.

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (8)$$

The symbol Σ represents summation, where y_i and x_i are the actual and calculated values, respectively, for the i th observation. N represents the total number of observations.

Root Mean Squared Error: Standard error of the residuals is represented by the Root Mean Square Error (RMSE) (prediction errors). Residuals are an estimate about how far away data points are from the regression line; RMSE is a metric of how wide out all these residuals are, this indicates how dense the information is just around the best fit line. To validate experimental data, root mean square error is often utilized in meteorology, forecast, and regression analysis.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2} \quad (9)$$

The symbol Σ represents summation, where y_i and x_i are the actual and calculated values, respectively, for the i th observation. N represents the total number of observations.

In the research, tried six different methods to find the machine-learning algorithm that would work best with the system. Although some of these methods actually use the same modeling, it would be best to evaluate them as a separate method since they do it with different libraries. In this way, this study aims to determine the most efficient model for the system.

3.2.1 ElasticNet regression

Linear regression tries to reveal this connection by considering that there is a linear connection between the target variables and the input variables.

In fact, this connection is expressed in the form of a line, if this connection is between a single input variable and the target, then the relationship between many input variables and target variables is a hyperplane, allowing them to be connected to each other. The coefficients of the model are determined using an optimization procedure aimed at minimizing the total squared error between the predictions (y_{that}) and the predicted target values (y). (Liu, Liang, Siegmund, & Lewinger, 2018)(Alhamzawi & Ali, 2018)(Z. Zhang et al., 2017)

$$loss = sum_i = \sum_0^n (y_i - y_{\text{that_}i})^2 \quad (10)$$

If the assumed coefficients of the model used are such that they are higher than the norm, this makes the model more sensitive to inputs while potentially making the model more unstable than normal. This phenomenon can be studied especially in equations with a small number of observations (sample) or more samples (n) than input estimators (p) or variables (called $p \gg n$ problems). (Zou & Hastie, 2003)

One of the best ways to examine the stability of the regression model is to modify the loss function to include an extra value for a model with large coefficients. Penalized linear

regression is used to describe linear regression models that use smoothed loss functions generated by the method suggested above during training.

Another common error method is to penalize a model in relation to the sum of the squared coefficient values. This is known as the L2 penalty. The L2 penalty reduces the size of all coefficients while preventing them from being removed from the model.

$$L2_penalty = \sum_{j=0}^p \beta_j^2 \quad (11)$$

In the equation, β_j refers to the j th coefficient or parameter in a model. It represents the weight or importance assigned to the corresponding feature or predictor variable. The term β_j^2 indicates that each coefficient is squared, and the sum of these squared coefficients is calculated to obtain the L2_penalty. Another common error method is to use the sum of the model's absolute coefficient values to penalize a linear model. This is commonly known as the L1 penalty. The largest difference of the L1 penalty from the L2 scale reduces the amount of all coefficients while allowing certain coefficients to be reduced to zero, removing the estimator from the model.(Hans, 2011)

$$L1_penalty = \sum_{j=0}^p \text{abs}(\beta_j) \quad (12)$$

Elasticnet is a penalized regression model that trains both with L1 and L2 penalties. Using terms in "The Elements of Statistical Learning," a hyperparameter "alpha" is specified to specify how many weights are given from each of the L1 and L2 penalties. The L1 penalty is weighted by a number between 0 and 1, and the L2 penalty is weighted by one minus the alpha value. The punishment mix is controlled by the parameter alpha, which is primary research conducted on qualitative grounds.

The benefit of using elasticnet is that it enables for a mixture of the both type of penalties, which can improve the performance of the fitting than a model with only one penalty on specific concerns.

Another hyperparameter, "lambda" is provided to alter the weighting of the sum of both penalties in relation to the loss function. By default, a value of 1.0 is being used to implement the completely weighted penalty; a quantity of zero eliminates the penalty. Lambda values as little as 1e-3 or less are common.

$$elastic_net_loss = loss + (lambda * elastic_net_penalty) \quad (13)$$

3.2.2 Multiple linear regression

Multiple linear regression is basically a statistical technique that tries to reveal the relationship between two or more independent variables and a dependent variable. You can usually use multiple linear regression to find:

The degree to which two or more independent variables are related to one dependent variable. (Andrews, 1974) The dependent variable's value at a given value of the independent variables. (Uyanık & Güler, 2013) All of the assumptions made by multiple linear regression are the same as those made by basic linear regression. Variance homogeneity (homoscedasticity); the amount of error in the estimation does not differ much between independent variable values.

Observation independence: The observations in the data set must be obtained using statistically acceptable techniques and it must be proven that there is no hidden connection between the variables. It should be noted that some independent variables in multiple linear regression might be related to each other; Therefore, it is very important for the health of the model to evaluate these relationships before constructing the regression model. If two independent variables are highly correlated ($r^2 > 0.6$), only one should be included in the regression model. (Majumdar, Naraseeyappa, & Ankalaki, 2017)

Normality: The data is distributed normally.

Linearity: For linear regression models, the optimal line between data points is a straight line, not a curve or some other type of grouping factor.

The formula for a multiple linear regression is:

$$y = B_0 + B_1X_1 + \dots + B_nX_n + e \quad (14)$$

In the given equation, y represents the predicted value of the dependent variable. B_0 represents the y-intercept, which represents the value of y when all parameters in the model, except B_0 , are set to 0. B_1X_1 represents the regression coefficient of the first independent variable (X_1), and it quantifies the effect that this variable has on the estimated y value when the value of X_1 is increased. Similarly, B_nX_n represents the regression coefficient of the last independent variable. The term e denotes the model error, which accounts for the amount of variation that remains unexplained in the y model.

Multiple linear regression aims to obtain the best-fit line for each independent variable by calculating the following three things:

- First, it takes the regression coefficients of the model, but these numbers are the ones that will give the least amount of total model error.
- Second, it calculates the p-value corresponding to the t-statistic of the whole model (this number shows how likely it is for the t-statistic to occur by chance if the null hypothesis of no relationship between the dependent and independent variables is true). The t-statistic and p-value for each coefficient of determination in the model are then computed.
- Third, the coefficient of determination (R-squared) is a statistical measure used to assess how much of an outcome's variance can be explained by changes in the independent variables. Even though the predictors are irrelevant to the outcome measure, adding more predictors to the Multiple linear regression model always increases R-squared.(Preacher, Curran, & Bauer, 2006) (X. Zheng et al., 2020)

As a result, R-squared cannot be used to identify which predictors should be included and which should be excluded from a model. R-squared will only be among 0 to 1, where 0 showing that neither of the independent factors can accurately predict the outcome and 1 suggesting that all of the independent factors can accurately predict the outcome.(Majumdar et al., 2017)

3.2.3 Bayesian regression

Bayesian linear regression provides a reasonably obvious technique for dealing with inadequate or poorly distributed data. It allows you to add priorities to coefficients and noise, which allows priorities to take over during a lack of information. More importantly, you can query Bayesian linear regression which aspects of the data fit and which are inherently unpredictable (you will

likely do this purely based on priorities). Thanks to the above-mentioned features, you have the chance to review:

What is the likelihood model connection, how sure could it be inside it, and what is the full likelihood function on it?

What is the predicted noise, and how does it differ from the posterior distribution?

What are the gradients and bayesian distribution estimates?

You may also investigate the direction and dispersion of the steepest climb with additional numerical effort. You may also inquire about the position of such an ideal or saddle-point and the probability on that (if doing quadratic regression).

Bayesian Linear Regression looks for the posterior distribution for the model parameters, while Linear Regression tries to find the single "best" value of the model parameters. In this way, instead of searching for the answer by a linear probability distribution as usual, this method also provides the model parameters to be produced by a distribution. The posterior probability of the model parameters depends on the training inputs and outputs:

$$P(B | y, X) = \frac{P(y | B, X) * P(B | X)}{P(y | X)} \quad (15)$$

The posterior probability density function of the hyperparameters provided the inputs and outputs is given by $P(y, X)$. This is equal to the data's likelihood, $P(y | X)$, multiplied by the parameters' prior probability, and split by a standardization constant. This is a straightforward formulation of Bayes Theorem, the essential premise of Bayesian Inference:

$$Posterior = \frac{Likelihood * Prior}{Normalization} \quad (16)$$

If to briefly explain what the formula means. You get a posterior distribution in hyperparameters which is proportional to the likelihood of a data multiplied by the prior probabilities of the parameters, as opposed to ordinary least squares. In this case, It can be seen two major advantages of Bayesian Linear Regression.

Priors: Unlike the frequentist technique, which assumes that all you understand about the variables comes from the data, it may add subject matter expertise or an estimate for whatever the parameter estimates should be if it have domain expertise or an estimate for whatever the model parameters ought to be. If you don't know any estimates ahead of time, you can use non-informative prior convictions for the variables, such as a normally distributed.(Muth, Oravecz, & Gabry, 2018)(Linero, 2018)(van Erp, Oberski, & Mulder, 2019)

Posterior: Bayesian Linear Regression produces a mixture of possible model parameters based on the data and the prior. This allows us to assess the uncertainty of our model: when you have less pieces of data, the posterior distribution may be more spread out.

The likelihood wipes out all the prior as the number of data points rises, so in the situation of unlimited data, the outcomes for the variables convergence to the obtained values via conventional least squares.(Pan, Qu, Liu, & Dias, 2020)(Haddad, Rahman, & Stedinger, 2012) (Koop & Poirier, 2004)

The depiction of design variables as distributions embodies the Bayesian worldview: it begins with a starting estimate, which is the previous estimate, and the built model gets less wrong as more data is acquired. Bayesian reasoning is a logical extension of intuition. It normally begins with a theory and then adjust its worldview as it gathers facts to support or refute that notion.

3.2.4 Polynomial regression

Polynomial regression is a subclass of regression model where the data is fitted with a polynomial function with a curvilinear link between the target attribute and the relationship between the independent variable. In a curvilinear relationship, the quantity of the target attribute fluctuates non-uniformly with respect to the predictor. In Linear Regression, it derives the following equation with a single predictor:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2^2 \quad (17)$$

In the regression equation, Y is the goal, X is the indicator, θ_0 is the biased, while θ_1 , θ_2 are the weights.

A linear connection can be represented by this linear equation.(Uyanık & Güler, 2013) (Kumar, Chinnam, & Tseng, 2019) However, a polynomial function of degree n is expressed in polynomial regression as:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2^2 + \dots + \theta_n X_n^n + \text{residual error} \quad (18)$$

Here, θ_0 represent the bias while $\theta_1, \theta_2, \dots, \theta_n$ are the weights within polynomial regression equation, and n seems to be the level of the polynomial. As the amount of n grows, so does the proportion of high terms, making the equation increasingly difficult.(Rezaie-Balf & Kisi, 2018) (Najafzadeh, Laucelli, & Zahiri, 2017)

3.2.5 Random forest regression

Random forest is a Supervised Machine learning technique that uses the group learning approach to do classification and regression. Random forest is a bagged technique, not a boosting technique. Trees in random forests grow in parallel. There is no contact between the trees while they are being built. It builds a huge number of decision trees during training and outputs the category that is the mode of the categories (classification) or mean forecast (regression) of the different trees.(Yuchi et al., 2019)

Random forest is built on the wisdom of crowds that posits that a huge proportion of statistically independent models functioning together as a panel will outperform any of the constituent models individually. This is because trees support each other by safeguarding one another from their own faults. A random forest does not interact between its individual trees. A random forest seems to be a strategy for estimating that mixes the outcomes of numerous decision trees to get the best possible outcome.(Mercadier & Lardy, 2019)(J. Zhang et al., 2019) A random forest is indeed a recursive (it aggregates the outcomes of numerous forecasts) that combines many decision trees with helpful adjustments.

The number of features that can be split at each node is limited to a specific percentage of the overall (which is known as the hyperparameter). This ensures that the network method is not overly reliant on any particular feature and that all possibly predictive variables are handled equitably. (Ishwaran & Lu, 2019)(G & Singh, 2017) When creating branches, every tree pulls

a representative selection from the raw data set, introducing an element of chance that avoids fitting problem.

The Random Forest Regressor manual lists a variety of parameters that may be used with the model. Several of the critical parameters are listed below:

- `n_estimators` — the number of decision trees to run inside the simulation criteria — the option allows you to choose the condition (loss function) that will be used to decide model outcomes. You can select among loss functions such as mean squared error (MSE) & mean absolute error (MAE). The preferred state is MSE.
- `max_depth` — determines each tree's maximum depth.
- `max_features` — the maximum amount of characteristics included by the model while calculating the divide.
- `bootstrap` — the possible values is Good, overall model follows the concepts of preloading.
- `max_samples` — This value indicates that preloading is enabled; otherwise, it has no effect. When true, this variable determines the greatest size of each sampling for each tree.

Synthesis learning is a method that combines the outcomes of many trained models along the similar data to get a more powerful predictive/classification outcome. The goal and requirement for supervised methods is to have every model's flaw (therefore in example, decision trees) to be autonomous and different from tree to tree.

Modularizing is the process of selecting subsets of a dataset at random over a predefined amount of cycles and attributes. Those outcomes are therefore aggregated to produce a more powerful result. An example of a various classifiers in action is bootstrapping.

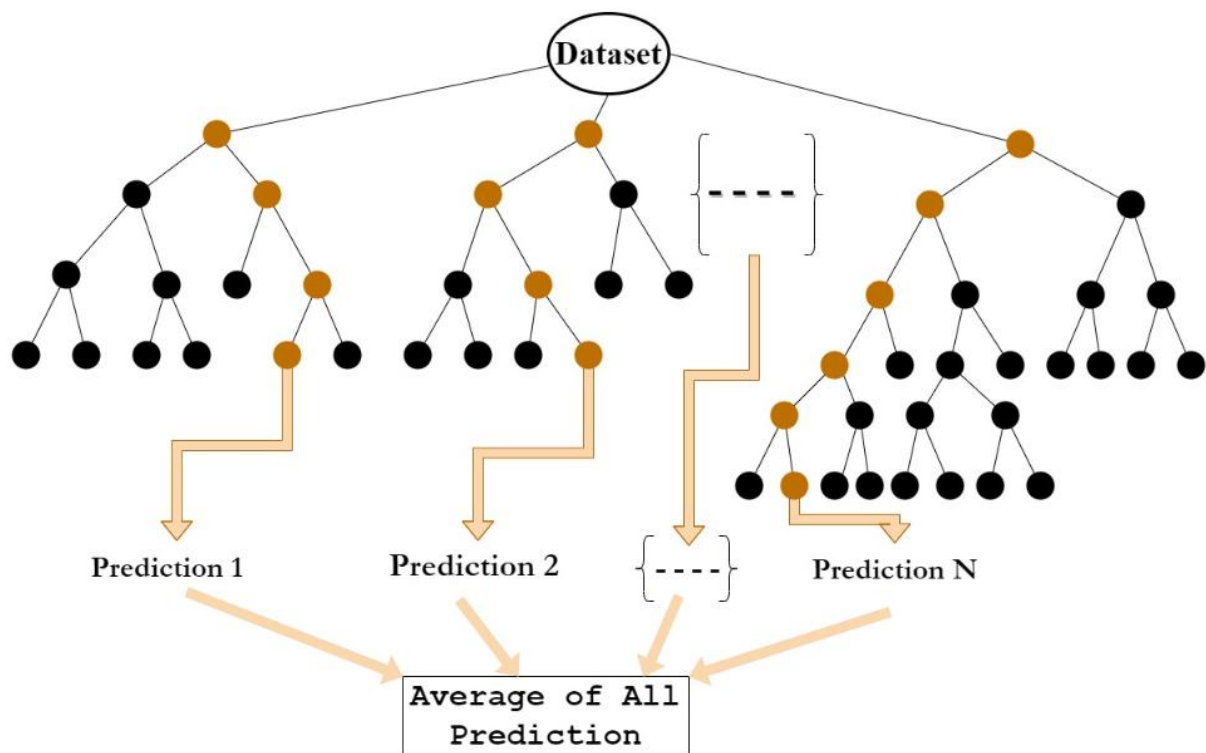


Figure 3. 8 Architecture of random forest algorithm

The above graphic (Figure 3.8) depicts the structure of a Random Forest Regression. Basically, trees run in parallel without any interaction between each other. During training, Random Forest creates many decision trees and estimates the average of the classes. Now you can go over the stages to have a greater comprehension of the Random Forest algorithm:

- From the training dataset, choose k pieces of data at random.
- Create a decision tree using the k pieces of information.
- Continue the first and second stages for the amount N of trees you wish to create.
- Create each one of the N -tree trees forecast the y - value for such new information topic under discussion and allocate it to the mean of all projected value of y .

A Random Forest Regression model is effective and precise. It often outperforms many problems, including those with non-linear connections. However, there really is no interpretability, overfitting is readily possible, and you must pick the amount of trees to include in the model.

3.2.6 Deep neural network regression

A neuron is the basic unit of the brain; the sympathetic nervous system has around 86 billion neurons that are linked by 10^{14} - 10^{15} synapses. After processing the signal, each neuron receives it off the synapses and sends it out. In order to build a neural network, this notion is borrowed from the brain.(T. He et al., 2020) Each neuron calculates the linear combination of the inputs and weights, applies biases, and emits the results. That whenever a large number of neurons collaborate to create a large number of outputs, a neural layer is formed. Finally, several layers come together to form a neural network.(Y. Xu, Du, Dai, & Lee, 2015)

Neural networks are produced when many neural layers unite to form a network, alternatively, it might claim that certain layers' outcomes become inputs for other levels. The completely connected layer is the most frequent form of layer used to build a basic neural network and conventions for naming, in which neighboring layers are fully linked pairwise and neurons in a single layer are not coupled to each other. In the N-layer neural network, we may not count the input layer. As a result, a mono neural net defined as a network with no hidden nodes (input directly mapped to output). In the case of the code used, it will utilize a single-layer neural network, which means there will be no hidden layer.(Bosse, Maniry, Müller, Wiegand, & Samek, 2018)(Achieng, 2019) The output layer. Unlike the neurons in the other layers of a Neural Network, the neurons in the output layer almost never have an activation function (or you can think of them as having a linear identity activation function).(Du, Tu, Dai, & Lee, 2016) This is due to the fact that the final output layer is characteristically used to represent arbitrary real-valued values or class scores (e.g., in classification), which is some kind of real-valued objective (e.g. in regression).It does not have an activation function because it is doing regression with a single layer.

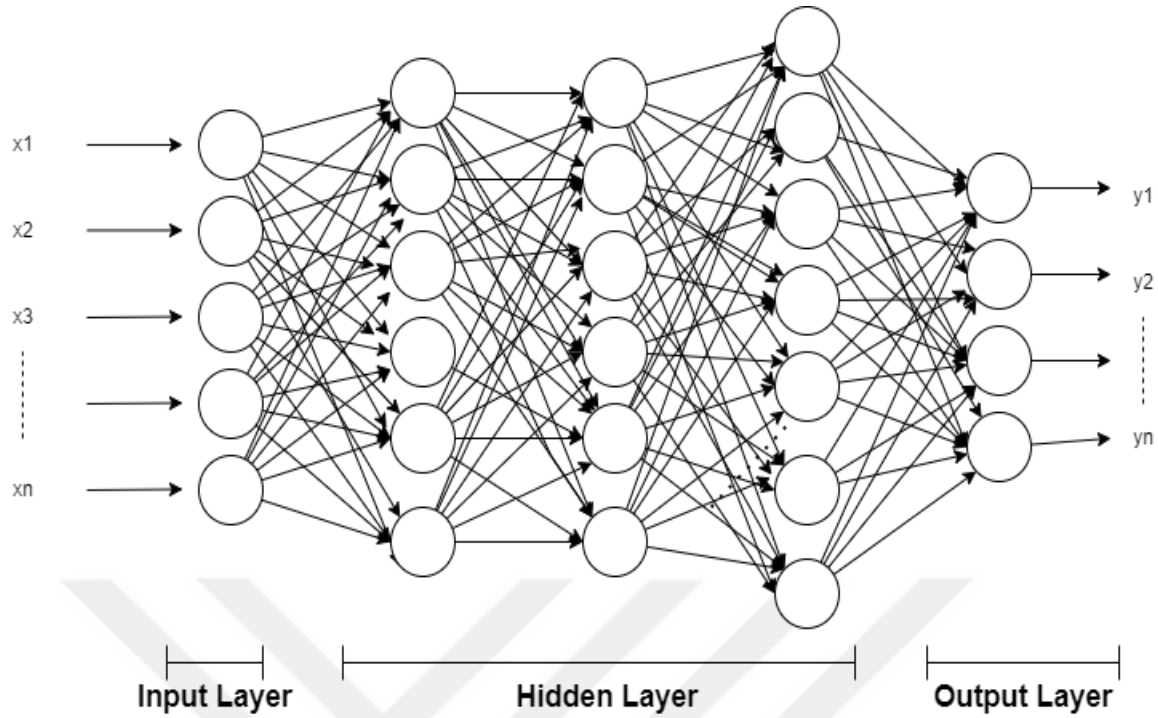


Figure 3. 9 Generic architecture of a Multi-Layer Perceptron.

Rosenblatt created the Perceptron, a neural network made of a single layer. The initial layer of units (or neurons) of the perceptron allows the user to "read" the data: each unit relates to each of the input parameters. Can be included a bias unit that is always turned on (1 is always transmitted). These units are linked to a single output unit, which gets the total of the units linked to it leasured by the connection weights, for p variables $x_1, x_2, \dots, x_p$; the output unit receives the starting weight (w_0) and the total of weights (w_1 to w_p) multiplied by respective variables (x_1 to x_p). The event is then activated by the output unit. Thus, a perceptron projected by the decision function f given by:

$$f(x) = \alpha(\sum_{j=1}^p w_j x_j + w_0) \quad (19)$$

It is not essential to transform the weighted sum obtained as an input in the case of a regression. The activation function is an identical function; it returns the sum of what has been shared. When training a perceptron, try to reduce the estimation error on the training set. As with the least-squares approach for linear regression, it could perform it directly. Neural networks are designed to be adaptable, which means that they change based on the information they receive.

Suppose that your n observations $x(1), x(2), \dots, x(n)$ are not observed concurrently, but rather sequentially, one after the other.

As a result, neural networks are taught using online learning algorithms, as contrast to the models you have seen thus far, which are learned using batch learning. Mini-batch learning is an intermediary solution in which the observations are conducted systematically. According to the above premise, it may infer that perceptron training is an incremental method. After each observation, the user will alter the weight of connections (w_1 to w_p) to lower the perceptron's prediction error at current moment (T_0). This work will use the gradient algorithm for solving it: The gradient indicates the direction of the function's more significant fluctuation (in this work it is the error function). The user must go in the reverse direction of the gradient to locate the minimum. (The gradient of the function is 0 when it is minimized locally.) The user will begin by randomly picking starting weight values for our connection ($w_0(0)$, $w_1(0)$, and $w_p(0)$). The user will then apply the new rule to each of the weights after each observation $x(i)$, $y(i)$. You can loop through the data set as many times as you like. People often iterate till the algorithm converges (the slope is sufficiently near to 0) or, more generally, for a set amount of iterations.

Hyperparameter (η) is a neural network element known as learning rate. Here are a few general guidelines:

If η is big, it is moving away from the optimal solution. If this had been already near to the ideal value, it may have exceeded its target and $w_j(t+1)$ been over its optimal. The algorithm might diverge, which means it will move away from the best result. Otherwise, if $w_j(t)$ is far from its ideal value and is small, the method will take a long time to converge. As a result, picking a learning pace is critical.

When it comes to regression, I choose the squared error. The following is the error function definition:

$$(f(x^{(i)}), y^{(i)}) = \frac{(y^{(i)} - f(x^{(i)}))^2}{2} \Leftrightarrow \frac{(y^{(i)} - \sum_{j=1}^p w_j x_p^{(i)})^2}{2} \quad (20)$$

Perceptrons are essentially linear models. His modeling power is inherently restricted, which caused some dissatisfaction in the research world in early 1970. Mathematician and educator Seymour Papert and cognitive scientist Marvin Minsky even cite their understanding of the futility of expanding the perceptron multi-layer in a classic 1969 book on the subject (Pollack, 1989).

The hidden layers in between input and output layers (see in the Figure 3.9). Each neuron in a layer is linked to every neuron in the layer above it. A multi-layer perceptron (MLP). Consider the case of a perceptron with a single hidden layer. The buried layer's neuron number h output z_h is derived by applying the neuron's activation function to a linear set of inputs:

$$z_h = a_h \left(\sum_{j=1}^p w_{hj} x_j \right) \quad (21)$$

The outcome then is obtained by applying the output neuron's activation function to the linear equation of the hidden layer's outputs:

$$f(x) = a \left(\sum_{h=1}^H v_h z_h \right) \Leftrightarrow a \left(\sum_{h=1}^H v_h a_h \left(\sum_{j=1}^p w_{hj} x_j \right) \right) \quad (22)$$

If the logistic function: $u / (1+e^{-u})$ was chosen as the input signal for all neurons, after which:

$$f(x) = \frac{1}{1 + \exp \left(- \sum_{h=1}^H v_h \frac{1}{1 + \exp \left(- \sum_{j=1}^p w_{hj} x_j \right)} \right)} \quad (23)$$

To boost efficiency is so often to use activation function, interlayer's function. The hyperbolic tangent function transforms a linear combination of various signals into a value between -1 and 1, rather than 0 and 1.

Cybenko's 1989 finding (improved by Hornik at 1991) (Hornik, 1991) states that any continuous function f specified on R may be estimated with reasonable precision by a neural network created by a single hidden layer if a significant number of neurons are used in that layer. This is a very positive outcome: you can now simulate any function! However, practice

has moderated this result: the number of neurons required in this hidden layer is, in many circumstances, gigantic, rendering neural networks including one hidden layer worthless. The answer is to build the hidden layers, resulting in what is today known as a deep neural network. This is the manifestation of deep learning. A multi-layer neural network may be understood by assuming that the output of each hidden layer is another method of representing data. The capacity of multi-layer neural networks to learn "excellent" representation is their strength. A representation between the final hidden layer and the output, such as a linear model (Perceptron), could be useful.

Backpropagation: Backpropagation is a gradient-based iterative method. This study will employ back-propagation, which was proposed for the first time in the 1960s and popularized thanks to Rumelhart, Hinton, and Williams in 1986 (Rumelhart, Hinton, & Williams, 1986). The idea is to split the mistake by using the chain rule theorem. The partial derivation of the error E depending on the weight w_{hj} may be written in the network at a hidden layer:

$$\frac{\partial E}{\partial w_{hj}} = \frac{\partial E}{\partial f(x)} * \frac{\partial f(x)}{\partial z_h} * \frac{\partial z_h}{\partial w_{hj}} \quad (24)$$

The following two terms are easy to compute since the logistic function's derivative is $u(u)(1-u)u'$:

$$\frac{\partial f(x)}{\partial z_h} = f(x) * (1 - f(x)) * v_h \quad (25)$$

$$\frac{\partial z_h}{\partial w_{hj}} = z_h (1 - z_h) * x_j \quad (26)$$

By calculating the partial derivative of the error with respect to the decision boundary and the partial derivatives of the outputs with regard to the preceding layers, the partial derivative of the error function with respect to a weight of a lower layer may be computed. Updating weights can thus be accomplished by alternating between a forward phase in which the intermediate layer outputs are updated and a backward phase in which the error gradient relates to the mass

of a layer can be determined by calculating from the gradient of the error function with respect to the mass of an upper layer. The weight is updated throughout the descent phase.

3.3 Docker Desktop

The main point of the digital twin concept is to monitor the solar panel system in real time and examine it in detail using a virtual model (Figure 3.10). In doing so, it has been simplified and aimed to monitor everything on a single platform instead of unnecessary data and large-sized complex systems. Thanks to Docker, you can examine all necessary platforms under one roof. While data and requests can be tracked on the platform with Redis, data from the real system or virtual model can be easily visualized with Grafana and you can monitor the containers with FastAPI, if necessary, you can manually upload data and test the system. All these data are stored with the created InfluxDB database.

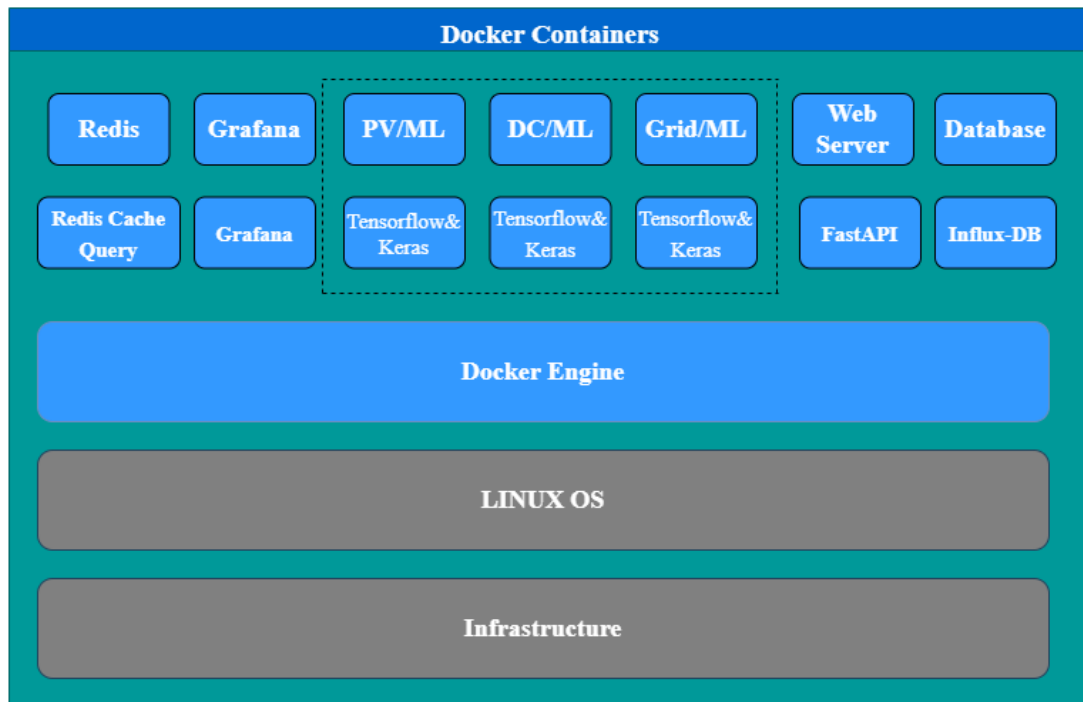


Figure 3. 10 The Docker Desktop schematic designed for this research

A container is a basic measure of application that wraps up software and all its components such that the program may be moved through one computer network to the other rapidly and effectively. The Container - based images are small, independent, operable software suite that

contains whatever necessary to execute an implementation: coding, os, operating systems, system modules, and configurations.

Container images become units at running time, and images become containers there in context of Docker whenever it executes on Docker Engine. Docker container software, which is accessible for both Windows and Linux-based apps, must always execute the same manner independently of the environment. Containers protect programs from its own eco system, ensuring that it runs consistently regardless variances such as improvement and testing.

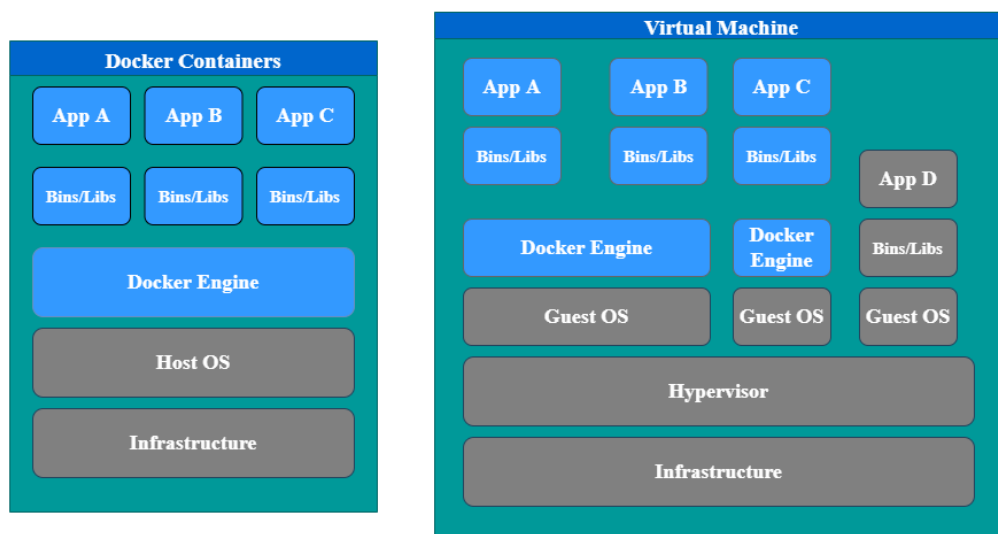


Figure 3. 11 The Docker Desktop schematic designed for this research

To understand the difference between Docker container and Virtual Machine, it is necessary to look at how they basically work:

If needs to think about why Docker Engine is preferred over virtual machine, it can show this with a comparison in the Figure 3.11. A container can be as small as thousands of bytes, but a virtual machine that has its own os might be numerous gigabytes. As a result, a dedicated server may support significantly more containers than virtual servers.

Some other significant advantage is that virtual machines might take numerous moments to wake it up the software packages and start executing the programs it serves, but container - based apps can launch virtually instantaneously. That implies containers may be launched "just

in time" once necessary then decommissioned while no longer necessary, reducing wastage for the owners.

A significant benefit of container shipping is how it enables for increased versatility. Instead of running a full sophisticated program within a single container, the project could be divided across parts (such as the database, the application front-end, etc). These are known as the microservices strategy. Due to every component is very basic, programs designed in this manner are simpler to handle, and modifications to sections may be done without needing to completely rebuild program. Individual elements (or subsystems) may be created just when required and are related items quickly due to the lightweight nature of containers.

A Docker container is a simulated default setting that is used in the development of new applications. It is used to develop, operate, and deliver software that is separate from the physical infrastructure. To execute multiple separate operations, a Docker container might use a single computer, exchange the kernels, and simulate the operating system. Therefore, Docker containers remain small.

3.3.1 InfluxDB

The data used in this study are in the form of time series, so it can be examined all changes in the system at any time. The main reason why our system is time-series is IoT devices and sensors. All these devices can generate a lot of data and InfluxDB is used as a database to cope with it. InfluxDB is a time series storage built to withstand heavy writing and querying workloads. This is an essential part of the TICK architecture. InfluxDB is intended being used as a backup database for any applications, which require significant volumes of time and date feeds, such as Continuous tracking, applications statistics, IoT sensor readings, and real-time analyzes.

InfluxDB is made up of measures that serve as a framework for tags, fields, and time columns. Each point is a collection of information contained in a line of an InfluxDB analysis. It could have two sorts of components in an unit of measure: fields and tags. Fields are made up of field keys and field variables. Tags, too, are made up of tag keys & tag values. Strings are used as field keys, while field values could be floats, strings, integers, or Booleans. Field values

represent the information, so influxDB cannot exist without them. Tags are saved as strings. They're both referenced and customizable. Tag array seems to be the filiation of tag public keys.

3.3.2 Redis cache

In this study, Redis is used as a Cache query; the reasons for this come from the basic features of Redis mentioned below. In addition, thanks to its simple language and popularity, it has a very large library and sample studies, and it has a structure that can be easily adapted according to the desired application.

Redis is a rapid in-memory storage and caches that is open source software underneath the License agreement, C based, and optimized for performance. Redis is an acronym for "REmote DIctionary Server".

Since its main data formats are comparable to its use in computer languages, such as characters, list, dictionaries (or hashes), groups, and having a series, Redis is frequently referred to as a database design platform. This also includes a plethora of different database systems and capabilities for approximation enumeration, localization, and parallel processing.

Redis data structures are the most similar to the native database systems used by developers in projects and computations of any NoSQL storage. Because core data architectures are easily distributed across operations and practices, it is great enabling fast evolution and speedy operations.

Redis stores information in ram by standard, with occasional drive permanence as the default. Because Redis maintains information on disk, it may be used as both a traditional relational database and a cache. Redis will deliver a mistake to a customer when it becomes overloaded, and it might be set as a cache to evict outdated and less meaningful information whenever new data becomes available.

Redis is typically preferred as a cache to hold frequently requested stored information, allowing programs to respond to clients. Redis provides a variety of clever caching strategies by allowing users to choose how longer users would like to hold data and which content to clear initially.

Smart caching serves numerous purposes, and has a significant influence on the consumer interactions, customer activity, website traffic, and retail revenue. For all these reasons, Redis cache is preferred in this study.

3.3.3 Grafana

Torkel degaard, chief architect, creator, and product manager, created Grafana in 2014. Degaard has attempted to make it simpler and more available to produce marker artworks by individuals by enabling it simpler and more straightforward to the items which other people find difficult to create. It is the open software panel and graphic detectable platform that provides you to enquire (for example, daily, measures, traces, and metrics), representation (visualized), a free and open - source panel and graphically detectable platform that enables you to perceive (Alert).

Grafana's advantages can be briefly listed as follows:

It manages to combine your data, not your dataset: if all of having a good data has become distributed and conflicting across platforms, or has gotten complicated for an inexperienced user, combining all of your information into one board, exploring your stacks, measurements, or events, and creating useful graphical representations provides.

Provides info that is visible to everyone: Groups and individuals that efficiently use data; Boards assist product managers, data analysts, and developers by providing access to common perspectives that keep everyone on the same page in order to spot problems, make targeted changes, and promote overall efficiency.

Panels that are simple to use for everyone: By constructing perspectives that everyone can comprehend the gathered metrics are shown as component sections on accessible panels.

3.3.4 FastAPI

FastApi's is used by the user of the created platforms to easily access. FastAPI is a web application framework that may be used to build contemporary and quick Python APIs.

Primary components:

Rapid: extremely rapidly, with performance comparable to NodeJS and Go, One among Python's quickest frameworks.

This is simple to program: It is around 200-300 percent quicker to design new functionalities.

* Fewer bugs: Programmers (human) mistakes are decreased by about 40%.

* Predictive text is ubiquitous. It can save hours troubleshooting.

Simple: In a method that is simple to learn and apply. The number of hours spent reading texts has decreased.

Brief: It has reduced source code and reduced the number of issues clients will experience through offering many options to declare the kinds of function arguments.



Figure 3. 12 The Web interface that the user will use to reach the digital twin

As can be seen in the Figure 3.12, our solar panel system is divided into 3 and it is possible to predict them separately. If the trained models cannot adapt to the solar panel systems after a certain period of time or if the system has been changed, there is a retraining mode to retrain with the data.

4. RESULTS AND DISCUSSIONS

In this section, going to share the results that is obtained as a result of training and using the machine learning models described above in this study, and will look at these data and try to determine the best method.

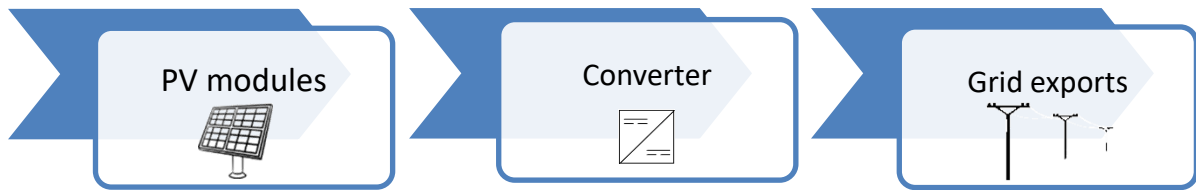


Figure 4. 1 It shows the whole system by dividing it into 3 partitions and states that it has been examined in this way here.

As shown in Figure 4.1, Solar Energy systems consist of too many complex parts and they work intertwined with each other, which makes it difficult to predict the system and to control and monitor it. With the proposed design in this study, it is aimed to reduce the complexity by examining the whole system in three parts separately from each other. On the other hand, as indicated by the arrows, the whole system is actually like a chain reaction, and it is desired to emphasize that there is a natural bond between them.

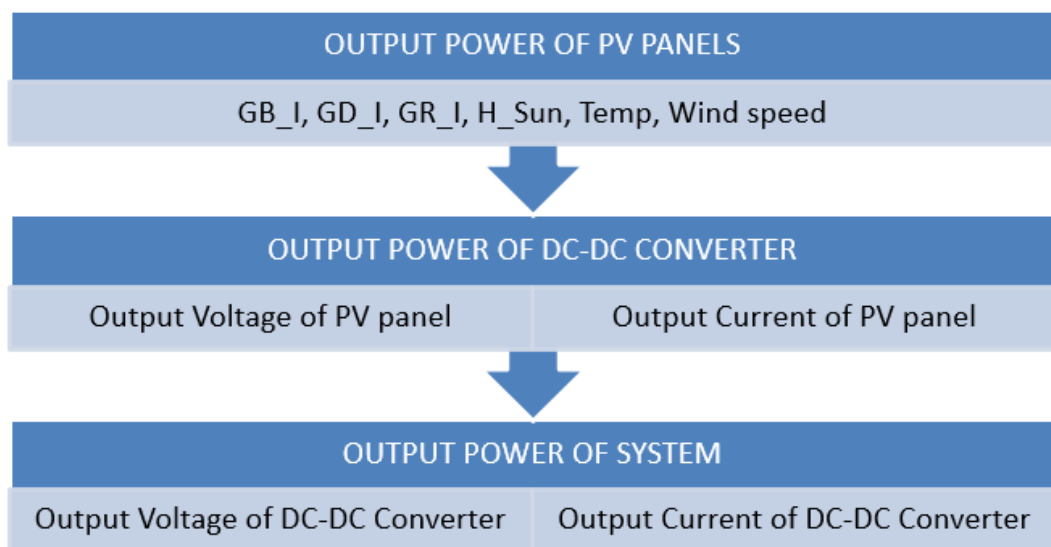


Figure 4. 2 It shows the input variables used to train Machine Learning algorithms and the outputs they are intended to predict.

In the Figure 4.2, the variables used for training the estimation algorithms and the variables to be estimated are shown. For the selection of all these data, solar panel systems have been examined and parameters that can show the characteristic structure of the system have been tried to be found. In addition, the connection of these variables with each other was examined in detail and the data structure was created in a way that would give the best results.

4.1 Pv panel part

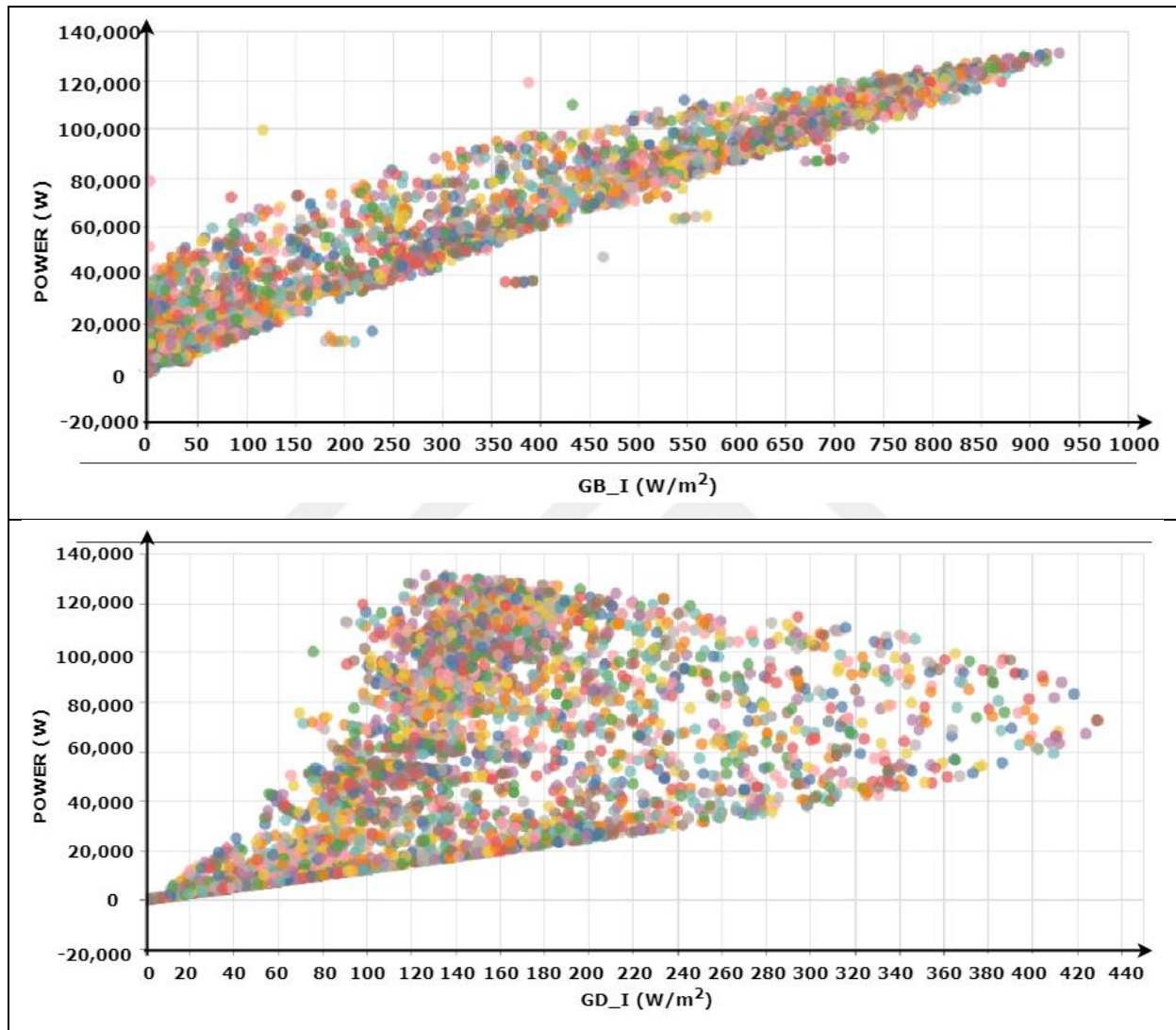
The solar panel part is the most important part of this study. It is the part that depends on the most variables, except where electricity generation takes place, and the input data needs to go through preprocessing to be estimated. In addition, since the production is completely dependent on the atmosphere and environmental effects, it must be constantly examined and protected.

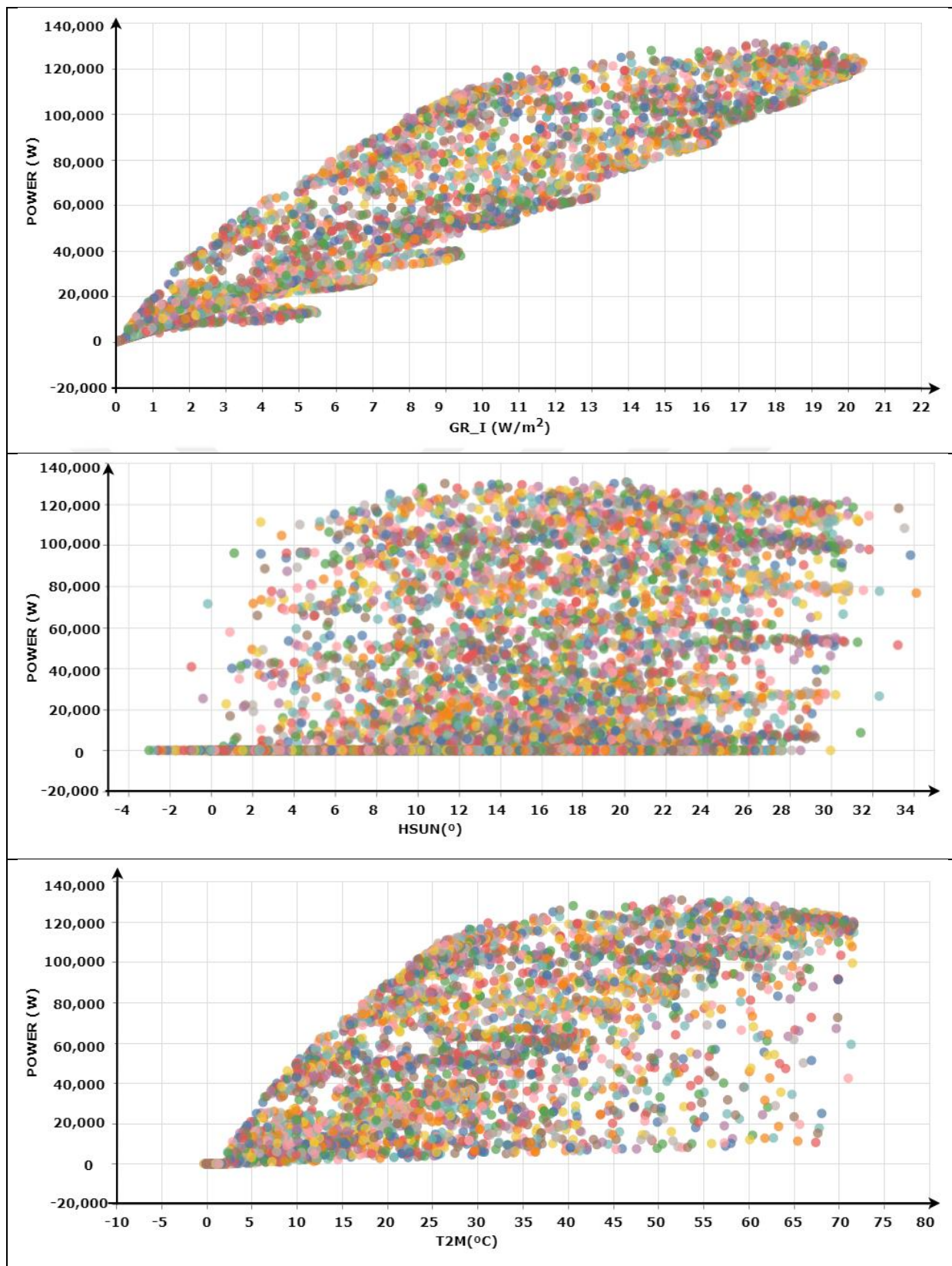


Figure 4.3 In this figure, you can see the correlation matrix of the variables between each other.

In this section, power estimation is tried to be made. In order to do this, the correlation of the above-mentioned variables with each other is examined. As can be seen in Figure 4.3 the first stage, different type of radiations and temperature are directly effective. However, it is stated that the wind speed (WS10m) and the temperature values above a certain altitude (H_sun) do not have a big effect. However, since it was determined that these variables had a certain effect

in our study, especially while training ANN; they continued to be used in the study. The main reasons for this are that the model makes predictions have more linear line characteristic and thanks to these values, it protects the model from overfitting.





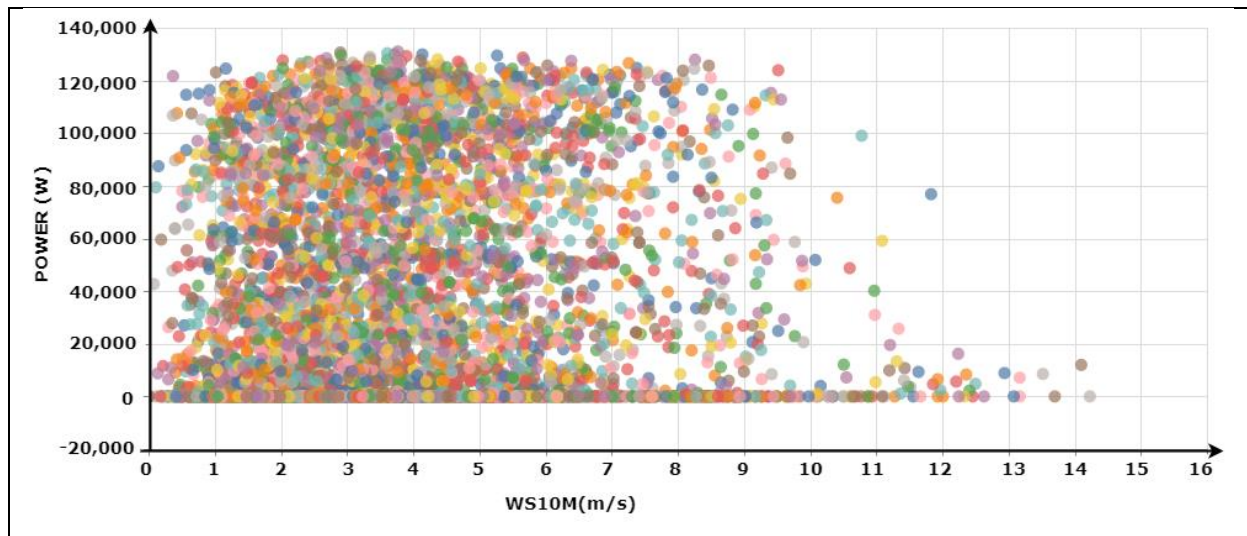


Figure 4. 4 In this figure, you can examine the connection between the produced power from the panel and the variables effect this production of panel. (Each dot shows the value of the energy produced on the panel and the other variable being measured at that moment)

Figure 4.4 visualizes the connection between the panel powers of the variables used for the PV Panel part. This image is very important in that it shows us the optimum range of values, especially for efficient production, but it should be noted that these variables are not the direct effect of these variables on power, because power is the result of a combination of them all. In Figure 4.3, it has already been studied using the correlation matrix for this purpose. Therefore, what is wanted to be shown in these figures is to capture the general correlation regarding electricity generation.

While Power is placed on the Y-axis, there are variables used on the X-axis. If you start by looking at the irradiancies, G_{b_i} and G_{r_i} basically behaved the same way. Along with this, G_{b_i} progresses in a more linear structure and it can be observed that the power values get closer to each other and increase as its value increases. While this linear increase is observed in the variable G_{r_i} , it can be argued that the range of variability of the power values is still high. The G_{d_i} variable exhibits a different character from the other two irradiancies and although it seems to increase in strength as its value increases, the power peaks between 100 and 230 values.

Looking at the H_{sun} variable, it can be seen that there is no direct link on power. When looked at this especially together with the fact that the power does not have a peak point, it can be seen that the power values are at the top and the bottom at the same level. The characteristic structure

of the T2m variable may seem familiar to us because it simply resembles the sigmoid function, and you can say that the power values generally become grouped at the top with the increase of this variable. This actually reveals the effect of heat on power, and it can also be observed that power generation stabilizes after a certain temperature value. The last variant going to look at, WS10m, exhibits a very different character from the others because the power is not grouped at any level and can be observed to be very dispersed. This inference can be supported by its value of 0.15 in the correlation matrix.

Table 4. 1 The names of the models used and their error values.

Name of Model/Parameter	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
ElasticNet Regrassion	8.59703481744653	104.91	10.24
Bayesian Linear Regrassion	7.99944116938373	836.02	28.91
3-degree polynomial Regression	1.02459138833258	5.38	2.32
2-degree Polynomial Regression	3.14059389173838	37.58	6.13
Linear Regression	8.10730610883819	93.98	9.69
Random Forest Regression	3.00672621793697	37.24	6.10
Deep Neural Network	0.04509104431161	0.02	0.06

If you look at the Table 4.1, you can easily see that the Deep Neural Network Regression model gives the best results at every stage. Predicting the production of solar panels highly depends on the input variables used and solving their complex interconnections and ratio proportions. In this regard, DNN and RF perform better than other models. Among the main reasons for this is that the two of them have a different architectural structure than the others, because the designs in the linear structure are clogged in intricate data after a certain level, however, RF and DNN, which are created through probability and trial and error base, are more dexterous in discovering these connections. After that, you can say that the 3rd degree Polynomial and Random Forest techniques come, respectively. What makes the Deep Neural Network technique so different

and more sensitive than the others is that it has a special mathematical formula that helps it reduce the previous error in each iteration, and that it can create a pattern between events that seem to be separate from each other due to its structure, making it easier to analyze the changes.

Backpropagation ANN is used in this study and this algorithm is explained mathematically in the previous section. In this part, the results of training with data will be shared. More than one model was tried during training. Especially, the number of layers and the neurons in these layers have been changed many times. In addition, changes were made in the learning rate and the number of epochs, and the results were monitored. After all these studies, the model took its final shape.

In this study, seven layers were used and five of them are hidden. Since the input layer is determined by the number of input variables, it contains 6 neurons, while the other layers consist of 256, 256, 512, 256 neurons, and finally, the result layer consists of a single neuron and that one gives us the power estimate. Rectified linear unit is preferred as the activation function, except for the result layer, and the linear activation function is used in the result layer. Dense layers consist of a Boolean class data structure. Adam optimization is a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments, and this was preferred in this study, mainly because it was thought to be suitable for our data, which has a non-linear structure (You can find the working schema in Annex-A). As a result of the studies, 500 was determined for the number of epoch and also the number of batches was adjusted to be 128.

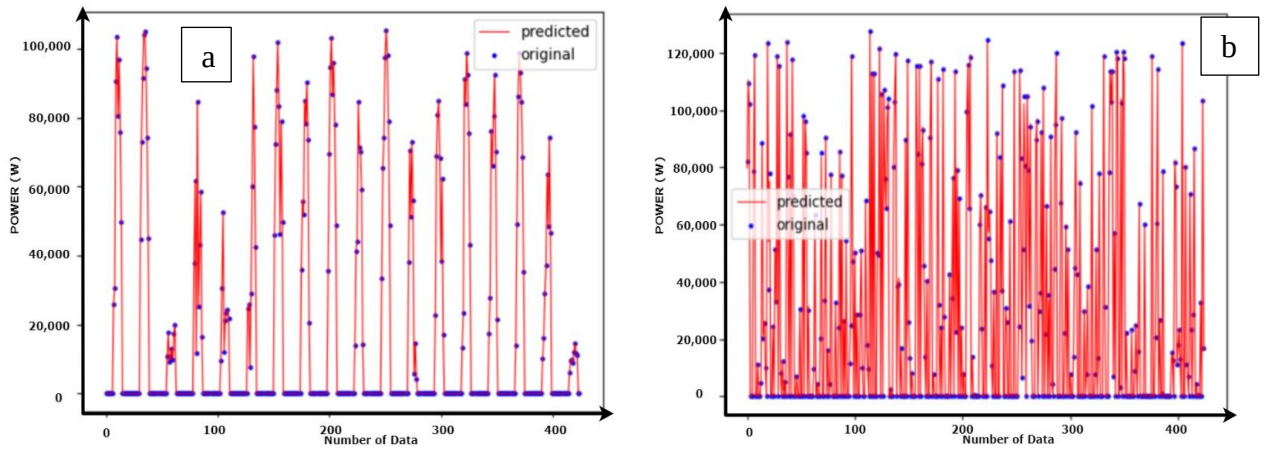


Figure 4. 5 The Comparison between the actual production results with the predictions made by the ML models **a)** Prediction performance of DNN Regression **b)** Prediction performance of Random Forest Regression

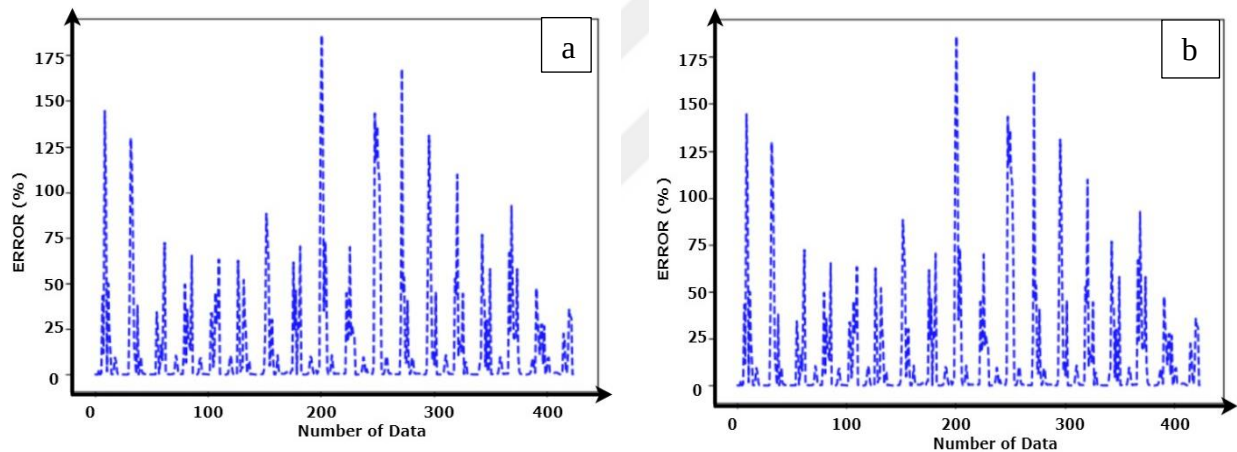


Figure 4. 6 Expression in Watts of the difference between the estimates made by ML models and the actual production values **a)** Error performance of DNN Regression **b)** Error performance of Random Forest Regression

As you can easily see from the figure 4.5, DNN is ahead of other models not only in error values but also in prediction. Figure 4.6 shows us the error rate in detail, as can be easily seen here; DNN regression is more sensitive than RF. The thing that makes the real difference is that DNN reacts very quickly against big changes. Therefore, DNN was preferred in this study. One of the reasons why DNN is preferred is that it is more flexible thanks to its architectural structure and that it can be done very easily if it needs to be retrained due to a change made in the system.

Without making any changes in the architecture, it will be sufficient to arrange only the loads and connections, so that extra costs can be avoided. It is also easier to work with the whole system because it can cooperate with the AI engine used in big data analysis. In addition, the trained model takes less space than RF model and can run on GPU or CPU, allowing users to work in different environments.

For the further examine you can check the (Annex-B) which shows the real, predicted value of the power produced by panels and the error differences also error rate. Error rate could look so high in some raw but if you check the rate with error differences; it is crystal clear that it is only high due to near zero power generation on the panels.

4.2 DC/DC converter part

Power electronic converters of various kinds are commonly used in distributed generation system operations, including regulated power supplies, power converters, DC power supplies, etc. AC/DC converters or rectifiers, DC/AC converters or inverters, AC/AC converters, DC/DC converters, and so on are examples of power electronic devices. Those converters are further categorized into many categories depending on various parameters. When it comes to DC/DC converters, there are three types: boost converters, buck converters, and buck-boost converters. In this section, goes over the DC/DC converters in depth, including their working principles and functions.

A DC/DC converter is an electronic systems device that receives a direct current input voltage and outputs a direct current voltage. The voltage output of the DC/DC converter might be larger than or less of the source voltage. These output voltages of the converters are utilized to fit the supply of power demanded by the workloads. A switch in the basic DC/DC inverter system may regulate its connections and disengagement of the power main supply. DC/DC inverter topologies are made up of a transistors or diode switch, energy storage systems such as inductances or capacitance, and thus are commonly employed as continuous voltage control or switched mode power supply.

DC/DC converters, unlike PV panels, have nonlinear behavior and are dynamic systems that can react very quickly. The main reason why the system is not linear at such a high level is that the semiconductor devices used in it are designed as key role. In addition, these switches used

in DC/DC converters produce nonlinear phenomena by causing parasitic capacitances and inductances in the system. Load fluctuations act as system disturbances in DC-DC converters. The output DC voltage of an uncontrolled electrical supply has a little ripple overlaid on a DC voltage that may then be eliminated with a voltage regulator. This regulator circuit is part of a larger DC/DC converter. A voltage divider for use in DC-DC conversion will reduce voltage however will still enable rippling to propagate to the exit until enough filtration is performed or the output is routed through a regulator.(Alagu et al., 2021) Since the data it has been used in this study consists of the data obtained from the measurement made every hour, the simulation results obtained from the DC/DC converter do not contain enough information. The main reasons for this can be listed as follows:

DC/DC converters are inherently non-linear, and this makes it very difficult for machine learning algorithms to capture the connections between them and express them with an equation.

In addition, DC/DC converters work very fast thanks to the semiconductor structures in them and instantly react and react to the changes in the system. This creates a big problem on systems like ours with long intervals between sample times. Because even when examining two consecutive data, the difference in DC/DC converter is huge and this poses a big problem for our machine learning algorithm.(Del Moral et al., 2015)(Martínez, Bolea, Grau, & Martínez, 2009) In order to overcome this problem, it has made a new simulation as you can see below (Figure 4.7). In this simulation, the DC converter is exactly the same as in our main system, and the change is in the input section. Thanks to the power generator used, it imitates a PV panel system with a single headlight, and thanks to this system, it can be monitored the changes in the system every 1 minute. In this way, it enables us to overcome the problems it encounters in the first and second items.

DC/DC converters are sensitive systems, and it is meaningless to examine the system only according to the power results, because the changes that may occur in the current or voltage directly affect the power and it needs to be examined in detail to understand which change occurs. That is why our machine learning algorithm consists of MIMO (multiple input multiple output) architecture. This change allows us to analyze the system better. The system estimates the current and voltage of the DC/DC converter as output.

Table 4. 2 The names of the models used and their error values.

Absolute Error	Mean Squared Error	Root Squared Error for Current	Root Squared Error for Voltage
Random Forest Regression	291.78502564	394017.5841	6627.70
Deep Neural Network Regression	10.29	3.21	3.008

As it was mentioned before, DC/DC Converters are non-linear in nature, so only Random Forrest and Deep Neural Network tests were performed in this section because, as explained in the previous section, other models are basically built on linear regression and there is no point in using them in this study. As it can clearly shows in Table 4.2. DNN is the algorithm that gives the best results. However, two separate error results for voltage and current are shown in this table, because even if the model is very good in overall performance due to having multiple variables, sometimes it can fail in one variable, and the best method to detect this is to examine each variable separately from each other in detail. The ANN machine learning used in this section is the same as used above, so it will not be mentioned again here. If you're curious about the schema of the model used here, you can check Annex-C. In this study, our model can say that it has developed a superior sensitivity to the voltage in the system and can say that it gives much better results in estimations by looking at the numbers. However, in order to examine it in more detail, it may be necessary to examine step by step with the help of graphics.

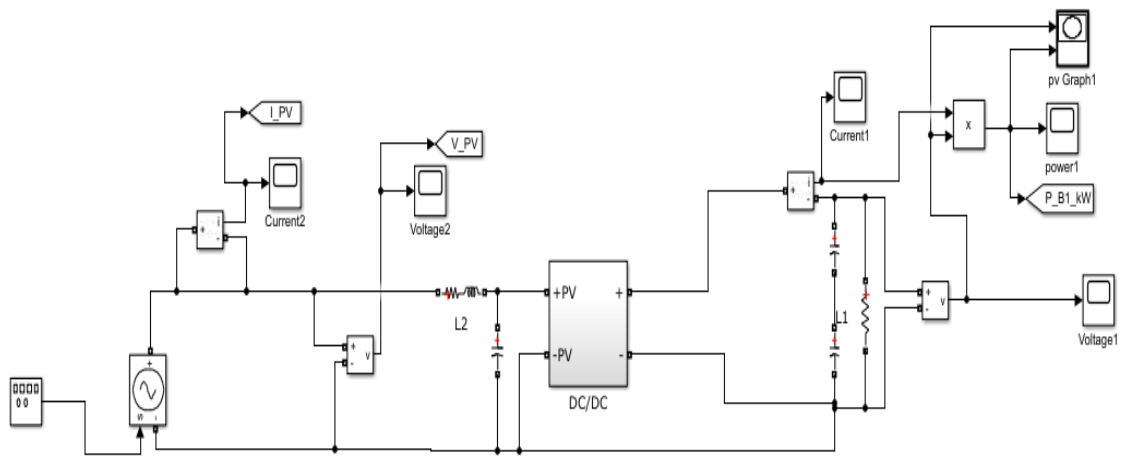


Figure 4. 7 Simulation of DC/DC Boost Converter using Controlled Voltage Source and Signal Generator as Power Source

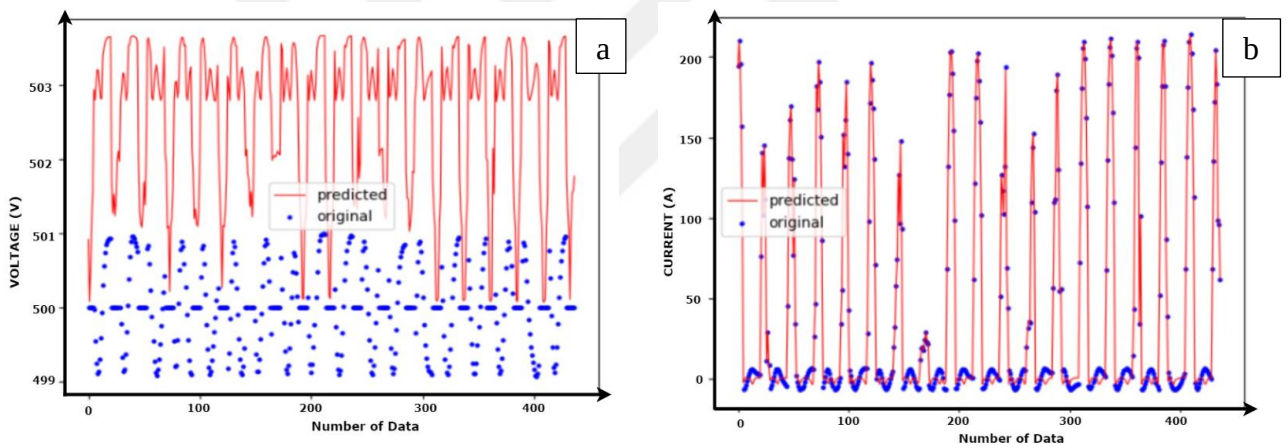


Figure 4. 8 Comparison between the estimates made by DNN Regression of the output voltage/current of the DC/DC converter and the actual results. **a.** the output voltage of DC/DC converter **b.** the output current of DC/DC converter.

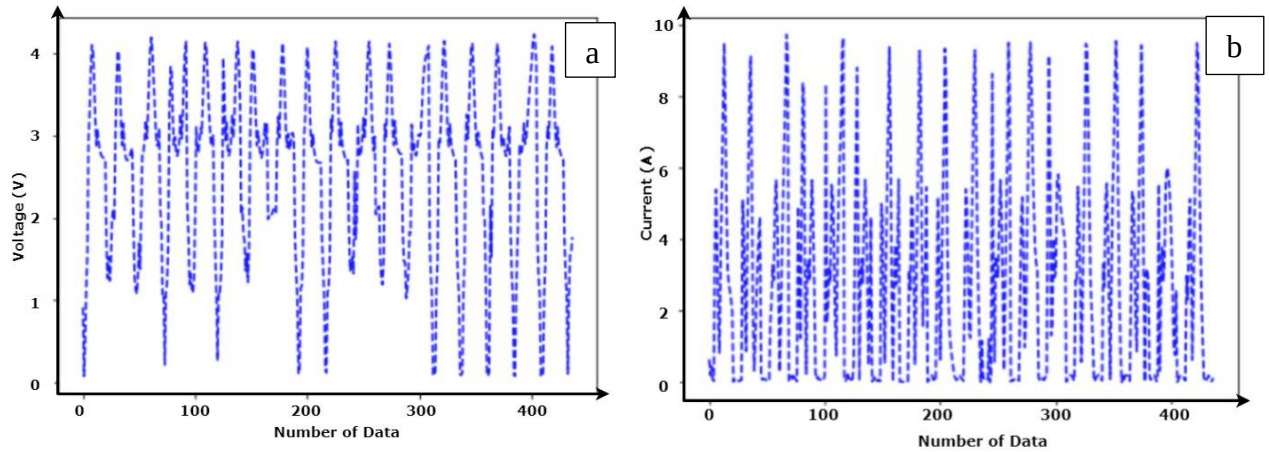


Figure 4.9 Expression of the difference between the predictions made by DNN Regression and the actual values in terms of voltage/Current **a.** Difference in voltage values **b.** Difference in current values.

Since the results in Table 4.2 show us that DNN is the most suitable system, there are only graphs of DNN in this section, so you will have a chance to examine in more detail.

If look at the figure 4.8 a), it can easily understand what exactly the values in the table mean. Our machine-learning program does a good job at estimating voltage, the highest difference between the estimation, the actual value is a difference of 4 volt, and the average error value is 2.5 volt, which is around 1% if looks at the average error rate.

If you look at figure 4.9 b), you can see that the current error rate is clearly higher than the voltage, which can be explained by the graphics in figure 4.8 because the voltage value generally continues in a more stable structure around 500 volts than the current, while the current is between 0 and 200 volts. You can see that it varies between amps, and this makes the job of machine learning difficult. However, if we look at figure 4.9 b) again, it can accept that the largest error value is close to 10 amps and the average error value is around 5 amps. It doesn't make much sense to make the error rate in this part because even if it assumes that the system is estimating 2 amps when the actual value is 1 amp - this would normally be a very good estimate in a situation where the variance is this high - it will appear to have 100% error, so the graphs It would be more logical to read through direct values.

For a more detailed examination, it shows the actual and estimated value of the Voltage and Current passing through the DC/DC Converter, and you can also refer to the error differences of these variables (Annex-D).

4.3 Power grid part

In order to examine the Grid section, three different situations were simulated, and the values obtained from these simulations were examined with the help of the DNN trained in the previous section. The first event includes the results for when our machine-learning model predicted the results of the simulated system using real values. In the second and third cases, it is used to detect the loss in the system due to the extra resistances connected to the cables going to the grid by machine learning and to show the damage caused by these losses in the system.

Table 4. 3 The names of the models used and their error values.

Name of Model/Parameter	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
ElasticNet Regrassion	2.24	1.50	1.020972
Bayesian Linear Regrassion	28.70	5.36	1.760091
3-degree polynomial Regression	1.57	1.25	0.91623733
2-degree Polynomial Regression	37.58	6.13	0.999668805938429
Linear Regression	3.29	1.81	1.3465
Random Forest Regression	0.0137	0.0017	0.045
Deep Neural Network	0.01	0.09	0.05376

When Table 4.3 is examined, it can be concluded that the grid part is easier to predict than the other two parts. Because all models performed relatively well. However, the two regression models have succeeded in distinguishing themselves from the others with their excellent performance. Since both of them give very close results, it would not be enough to choose the

model to be made in this section only according to the error results. Therefore, as mentioned above, a test setup was prepared, and the results were analyzed in detail with graphics below.

Backpropagation ANN is used again as an algorithm in this section. In addition, the dense layer containing 512 neurons of the model used in the first section was removed. The reason for removing this layer is that, as mentioned above, the data structure can be easily estimated due to its homogeneity. By removing this layer, there is a chance to speed up the algorithm as well as to reduce the possibility of overfitting. All other hyperparameters are used to remain the same (You can find the architectural structure of the ANN for grid part in Annex-E).

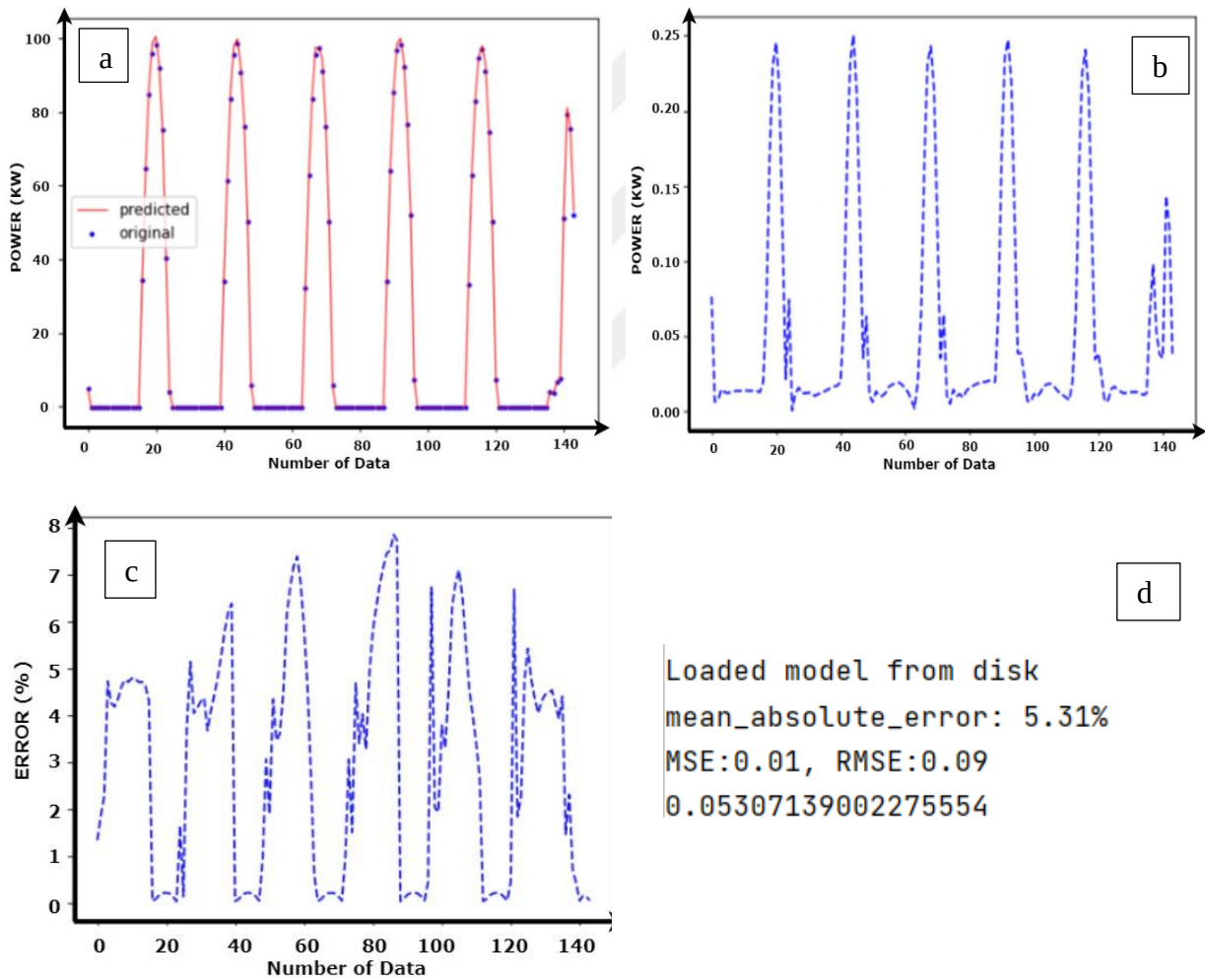


Figure 4. 10 The results obtained from the data used for testing purposes and has not been tampered with **a)** It shows the estimation and the actual KiloWatts (KW) values on the same graph **b.)** It shows the difference between the estimated and actual Watts values in terms of KiloWatts (KW) **c.)** It shows error rate between the estimated and actual Watts values in terms of KiloWatts (KW) **d.)** It shows error values in terms of mathematical calculations.

Looking at the raw results (Figure 4.10), the RMSE rate is only 0.09% and the Error rate is around 5.31% in average, these graphs you can see how high precision and accuracy the model was trained works with. It is possible to examine that the difference between the actual data and the predicted is very small.

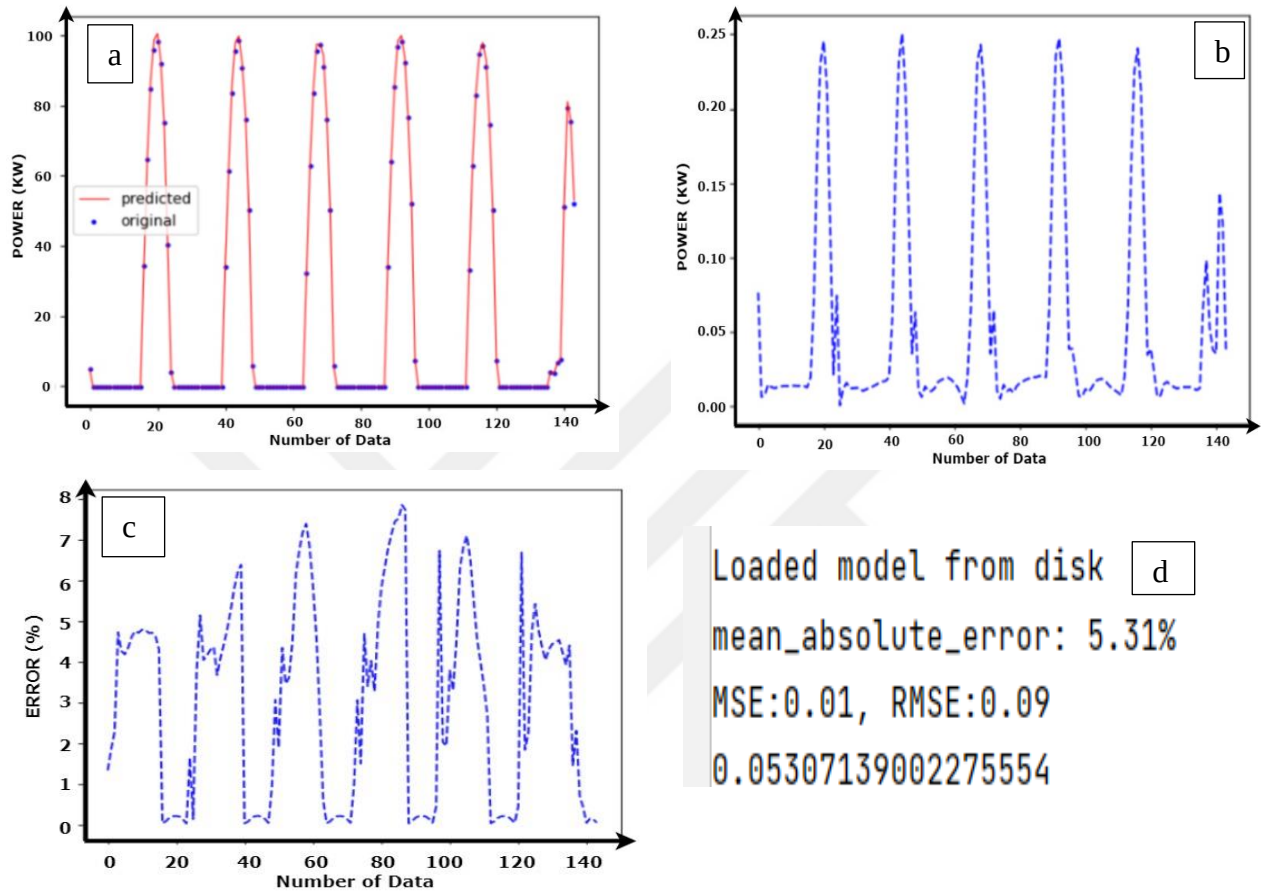


Figure 4. 11 The results obtained from the system connected to the 100-ohm (resistance, capacitor and inductor) connected in series to the grid part in the simulation to be used for testing purposes. **a.)** It shows the estimation and the actual KiloWatts (KW) values on the same graph **b.)** It shows the difference between the estimated and actual Watts values in terms of KiloWatts (KW) **c.)** It shows error rate between the estimated and actual Watts values in terms of KiloWatts (KW) **d.)** It shows error values in terms of mathematical calculations.

The graphs in Figures 4.11 and 4.12 are used to represent the losses in the system due to the extra resistors added on the grid line. If you can examine the mathematical error rates by looking at the figure d's, it can make it easier for us to understand what it means visually. While the MAE ratio is 30% when the resistance is 100 ohms, you can see that this ratio increases to 286%

when the resistance is increased to 1000-ohms. It can also be seen from the large change in the energy loss experienced with the increase in resistance from the RMSE ratio. These numbers tell us how accurately our machine-learning model works.

Let's start examining the graphs from a, because it can see the difference between the predicted and the real at the same time, and if the loss is too much, it can easily examine it. If it looks at the results on the unmodified data (Figure 4.10), it can be observed how well both works together and the model works very well. If looked at the results manipulated with a 100-ohm resistor (Figure 4.11), see that the difference is not very large. Finally, if it examines the results manipulated with a 1000-ohm resistor (Figure 4.12), can be seen difference between the data that should normally be, and the value lost due to the resistance.

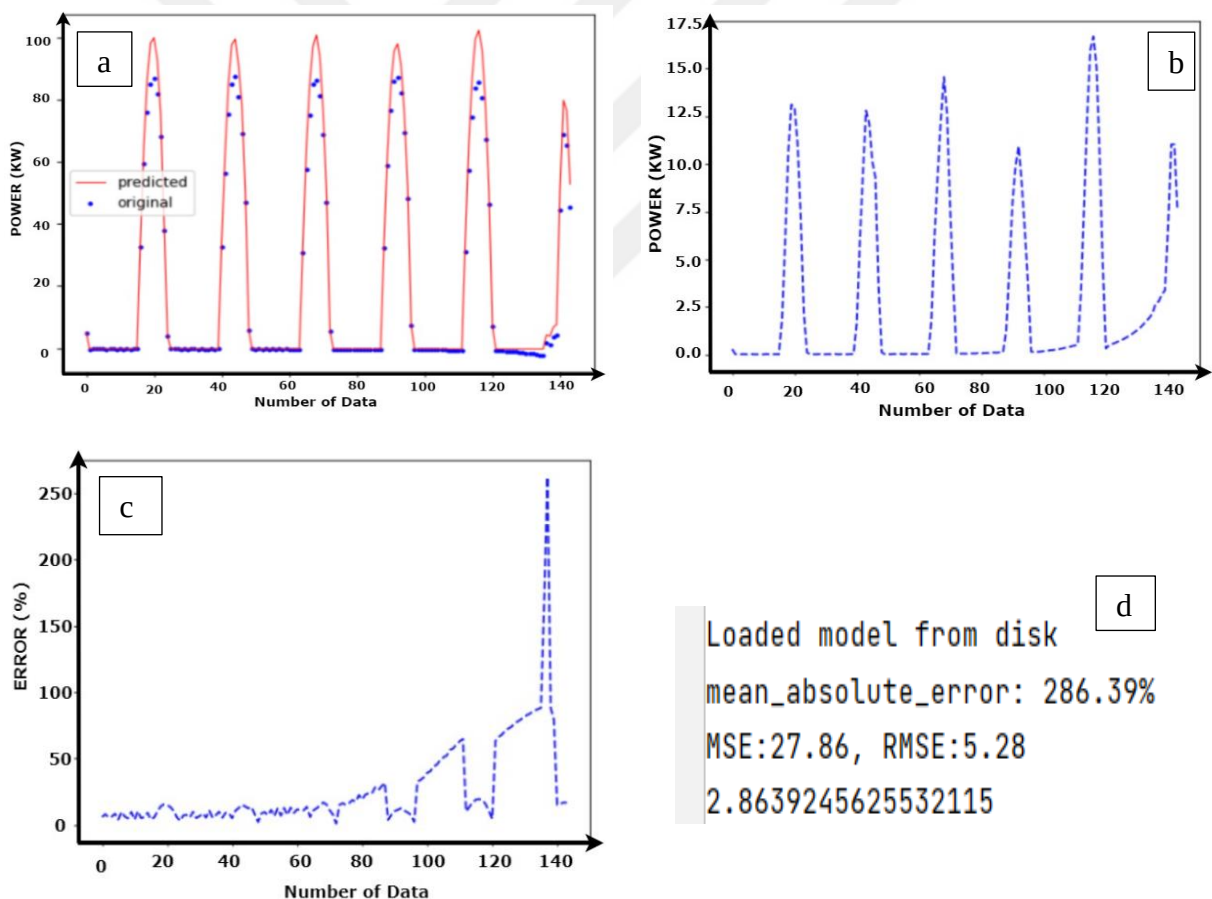


Figure 4. 12 The results obtained from the system connected to the 1000-ohm (resistance, capacitor and inductor) connected in series to the grid part in the simulation to be used for testing purposes. **a)** It shows the estimation and the actual KiloWatts (KW) values on the same graph **b.)** It shows the difference between the estimated and actual Watts values in terms of KiloWatts (KW) **c.)** It shows error rate between the estimated and actual Watts values in terms of KiloWatts (KW) **d.)** It shows error values in terms of mathematical calculations.

In figure 4.12 b), it can see the errors directly in KiloWatts, and in figure 4.12 c), you can see the equivalent of this error as a ratio. It should be kept in mind that if the actual value is close to zero, even if the error difference is very small, the rate may be very high, so only the error rate should be considered in graph d.

For a more detailed analysis, you can review Annex-F, which shows the actual and estimated value of the power in the network, error differences and error rate. In this section, the error rate is not very high, especially if you compare it with the PV panel part, but if check it with error rate differences and error value differences, you can see that the algorithm works very well.

4.4 Cleaning optimizer

Over the six-year research period, 75% of the days had a negative soiling rate (i.e., an increase in soiling on the PV panels), whereas the remainder 25% had a particular benefit due to the resuspension of dust from the panels caused by the high wind. Key parameters influencing seasonal soiling variations, as well as their corresponding connections, were also discussed. It was discovered that a minimal rainfall of 3 mm was all that was needed to completely clean the Photovoltaic panels. Dust storms were more common during the summer season while soiling damages were highest, but the influence of sandstorms during the winter season was shown to be extra significant compared to summer.

Due to predicting meteorological data for a single year is extremely challenging, it will rely on statistical approaches to provide estimates. It will be important to have as extensive a history as feasible (minimum of one year) for such deployment of irradiation and ambient temperature at the hourly level, in addition to rainfall on a daily basis, where is wanted to make the forecast. Because SuperPV demo sites do not have a long history, this information must be obtained from an outside source. Based on this information, the mean of every metric one per daily or hourly slot on a specified day in the history will be generated for the various years for which the information is available.

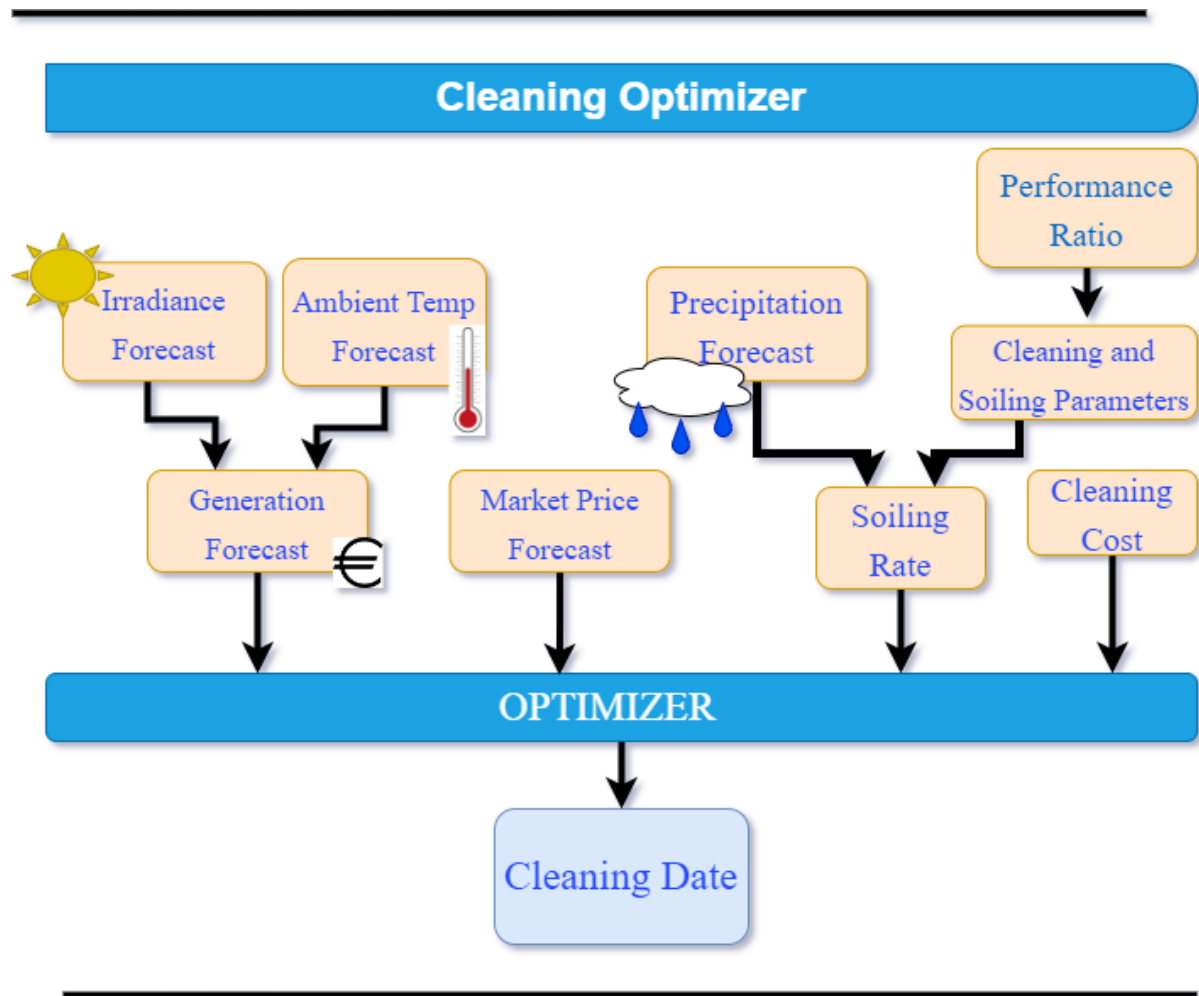


Figure 4. 13 Cleaning optimizer diagram designed for this project.

The generation preliminary analyses will be trained in the same manner as stated in the preceding part of Digital Twin, but with the ambient temperature taken into consideration, as the screen heat is more challenging to get long histories. If a temperature history for the panel exists, it might be used. This algorithm can be used to forecast installation generation in hourly time windows over the following 365 days.

Because power market prices, like meteorological data, are very unpredictable, will use a statistical estimate in the lack of more dependable instruments. For this, it will need at least one year's worth of energy costs to compute the expected price. The forecast will be generated by averaging the past years, taking into consideration the same day of the week nearest to the day beneath research in prior years.

It will employ two characteristics to determine how unclean the plant is on days when there is no rainfall and how the plant would be cleansed when it rains. It uses the knowledge that rainfall on solar panels cleans the panels, but a lack of rain or human cleaning results in dirt collection, making the panels less efficient. To investigate the soiling behavior of the installations, it will be essential to compute the Performance Metric (which is the power prediction and is done in other portions) and the daily rainfall.

For the soiling parameter, it constructed the Theil-Sen estimate during intervals of greater than 14 days without precipitation. This will provide us with the slope at which the Performance Metric is deteriorating. The soiling parameter will be the average of all dry seasons. In the instance of the rain-induced self-cleaning of installations, we take rainfall periods and evaluate the Performance Metric of the morning before the rain begins with the day following.

In the instance of the research of self-cleaning of facilities when it rains, it chooses rainy intervals and correlate the Performance Metric of the day before the rains come with the day following, demonstrating how performance increases with the influence of rain. It shall derive two parameters from this analysis:

Minimum Precipitation: It has been noticed that a modest amount of rainfall does not affect performance enhancement.

Cleanliness per mm: It will determine the percentage of increased Performance Metric divided by mm precipitating to achieve the metric of progress per mm produced.

For each of the 365 days to be examined by the optimizer, a vector will be created using the Soiling Rate Differential. Figure 4.14 represents the volatility of the plant's production depending on the level of soiling and the amount of rainfall generated.

For this computation, the following procedures will be taken:

1. Beginning with the daily precipitations, the accumulated precipitation on successive rainy days will be calculated.
2. The following reasoning will be used to compute the vector:
 - a. When there is no precipitation, the result of Soiling Rate Slope will be negative.
 - b. For precipitation larger than the minimal precipitation, the cleaning indication per mm multiplied by the quantity of rainwater generated will rise.
 - d. The value shall be 0 for rainfall below the minimum precipitation.

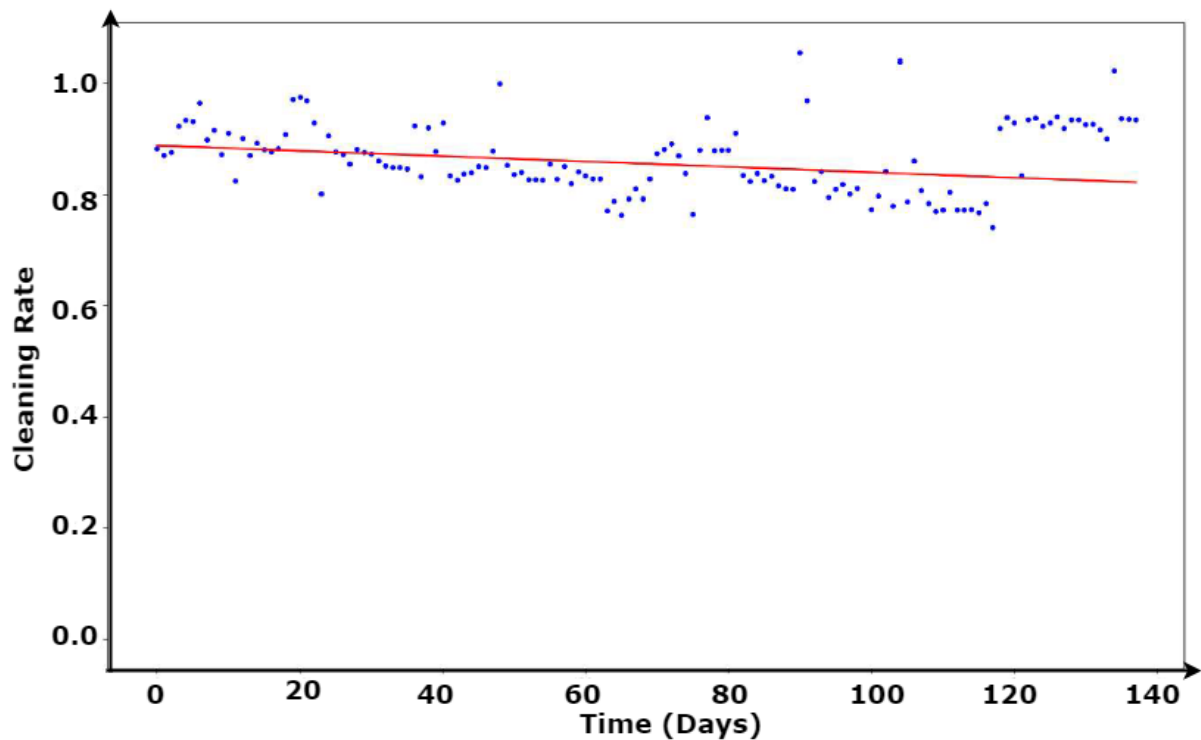


Figure 4. 14 Soiling rate slope

If mechanical cleanup of the facility is necessary, an approximate cost will be quoted. This will allow users to compare the cost of manual cleaning against the return on electricity generation. It gives an optimization issue in which the plant's profit from energy sales for the next 365 days is maximized based on when it is most feasible to clean the facility. It will utilize Python's Pyomo package for this. Pyomo is an open-source toolkit for creating optimization models. The CBC solver, that is also open source, will be used. It goes through the following aspects that will be utilized in the optimization issue, some of which have already been discussed.

Parameters are vectors that should be given in order to solve the issue.

Market Price: A vector of energy costs for every hour of the year.

Cleaning Cost: A vector representing the cost of cleaning the plant. Any day throughout the year will have 365 equal values.

PV Generation: Forecasted hourly generation over the next 365 days.

Soiling Rate Differential (SRD): The oscillation in the Soiling Rate caused by the rainfall prediction.

Constraint: the rules that must be followed in order to complete the optimization.

Loss Dirty: This is an equation used to quantify the proportion of losses caused by soiling. It is the total of the previous day's Loss Dirty plus the current day's Soiling Rate Differential, with the clean applied to begin the dirtying if washing is performed.

$$loss_dirty_i \geq loss_dirty_i + srd_n - clean_i \text{ for } i \in 1, \dots, 365 \quad (27)$$

$$loss_dirty_i \leq 1 \quad (28)$$

Variables are the vectors which will be generated when the issue is solved:

Clean Days: This is a 365-day vector that indicates when the plant needs to be cleaned.

Objective: is a function that produces the value to be optimized.

$$\max \sum_{i=1}^{365} \left(\sum_{n=7}^{24} (price_n * (gen_n - loss_dirty_i * gen_n) - cost_i * clean_i) \right) \quad (29)$$

The goal with this equation is to continue to ensure that it is economically profitable while calculating the cost of cleaning.

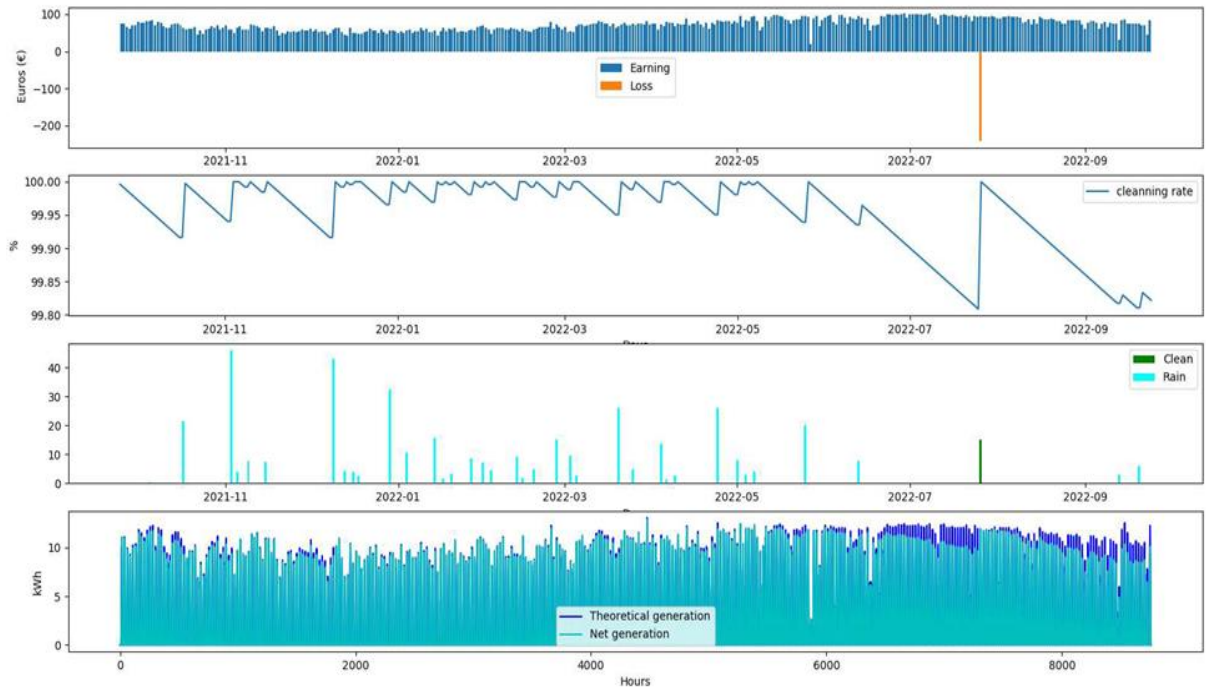


Figure 4. 15 Dashboard in the grafana show the cleaning parameters and results.

In the Figure 4.15, first graph shows the comparison of the profits produced by the sale of energy and the losses by the cost of cleaning. It shows that the general benefit during the 365 days is € 559.54 if you check it detailly. In the second graph observes the percentage of dirt on the panels during the 365 days under study, you can compare it with the third graph where the days where rainfall is predicted, and it is shown where it is proposed to clean the plant to maximize the economic benefit. Finally, there is a comparison of generation under ideal cleaning conditions (Theoretical generation) and generation with calculated dirt conditions (Net generation).

4.5 Monitoring and evaluation part

You see the grid part in the figure 4.16 and figure 4.17, thanks to Grafana, can examine the system in real time. The graph you see at the top actually shows the prediction of machine learning with electricity passing through the grid, you can see this in detail in the next figure. The dashboard in the middle shows the error rate between the estimated and actual power, as an example in this study, 20% was considered tolerable, and if it exceeds this rate, the system sends an error to the user. At the bottom, you see the types of irradiations, where the data is obtained and sent to the system thanks to IoT devices. This part can be supported with more sensors to facilitate the monitoring of the system.

One of the output goals of the Digital Twin is to create a digital twin of the physical system and connect it to the real one to create a system that is easy to examine, to ensure a seamless connection between the IoT and data analytics. The Digital Twin environment provides rapid assessment and real-time decisions based on reliable data. Digital Twin is one of the concepts that will be most suitable for the modeling and data set has. It has a working principle that it can create the same system in digital form as a structure and can be summarized as copying its behavior under normal conditions. In this way, it can easily detect when there is any energy loss in the running system or when there is a power outage caused by structural problems without wearing out the components. Considering that the important thing in today's world is time and efficiency; this gives us a chance to intervene before the fault is completely completed. However, it is clear that the current system and concept is supported by sensors and a larger and continuous dataset is needed to achieve the goal mentioned here.



Figure 4. 16 Dashboard in grafana with model input data with grid part



Figure 4. 17 Examination of a detailed part of the model input data with the energy results of the grid section

5. CONCLUSIONS

It will be very difficult to control the entire system with a single machine learning and instantaneously examine all inputs and outputs and keep efficiency at the highest level compared to past productivity. The reason for this is that solar panel systems are generally non-linear and produce instantaneous response in the tools they contain and generate noise in the system. To prevent all of this, the system is imaginatively divided into 3 parts. Thanks to this fragmentation, users can monitor the entire system instantly and make it easier to measure the region and its effect on the system in case of any failure. You can clearly see this from the results above. In this study, six different algorithms were used - you can see details of the models and their parameters at the Annex- G - and the results obtained showed that the Deep Neural Network Regression algorithm was the most suitable. DNN makes it easy to work with non-linear systems, thanks to its very fast response structure and algorithm.

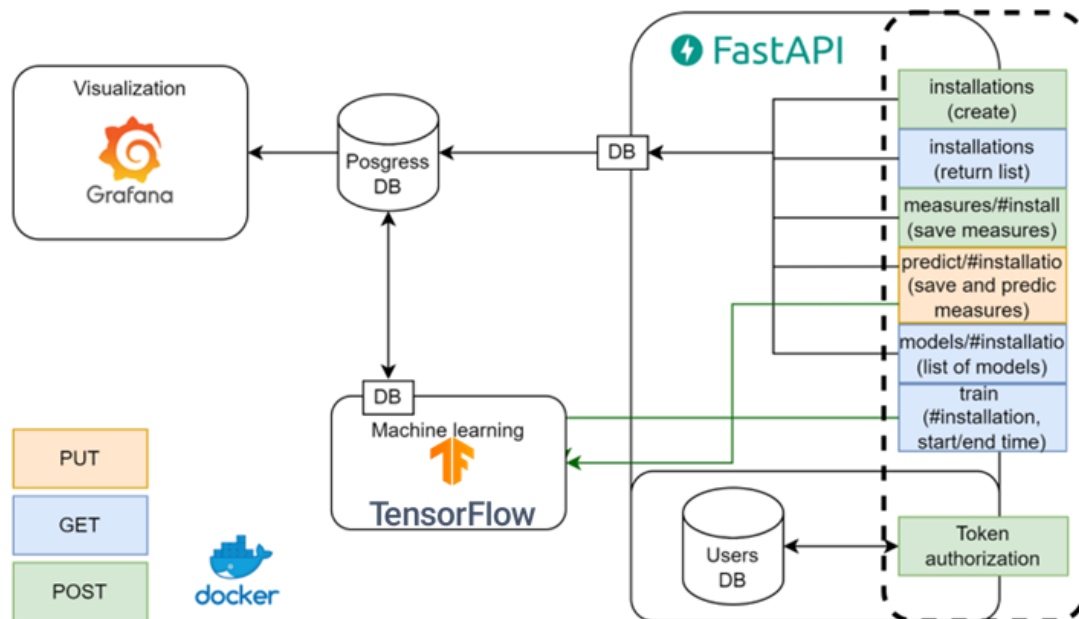


Figure 5. 1 Digital Twin application prepared for this research.

This figure 5.1 shows the first part of the Digital Twin concept, which creates a virtual version of the PV Plan and uses real-time data to compare the efficiency of our physical plant. In addition, if any changes have been made to the physical system or there has been a big change in the environmental conditions, you can retrain the system with all these new variables to

synchronize it with the virtual twin. In this way, it is very important to accurately estimate the efficiency of the solar panel, which has decreased over the years, and to follow the estimated electricity production capacity of the system also designed a cleaning optimization for this project to keep the plant efficient. Thanks to the local database created specifically for the users, it is easier to make a retrospective analysis of the system, and thanks to grafana, all this data is prepared in detail and its analysis is facilitated. This study has reached the following results:

- Since there are too many nonlinear parts in a solar power plant, it is more difficult to follow the system.
- By dividing the system into three ends, the complexity and noise in the system have been reduced and the system monitoring has been increased.
- Multiple machine learning methods have been tested and DNN has given the best results each time.
- Docker Desktop has been used so that DT can be used in all operating systems. The aim here is to have the chance to use all sub-platforms together.
- With cleaning optimization, it has a chance to reduce the unstable structure of the system while avoiding unnecessary expense by cleaning only when necessary.
- Since the data structure is time series, grafana is used to compare and visualize the results.
- Although the incoming data in this study was obtained from IoT tools, a separate tracking system was not used for them, only the results were recorded in a database.
- A web-based platform is used to allow the system to react quickly and to reduce complexity.

6. RECOMMENDATIONS

In today's technology, IoT and sensors are still not cheap and accessible enough; every innovation and progress in these areas in the coming years will create an important opportunity for the digitization of solar panel systems, along with the necessary regulations for the existing technologies to work harmoniously and the strengthening of global production and logistics networks. They need to be supported by countries and international institutions for this purpose.

Grid operators must securely reset the whole system if a catastrophic natural catastrophe or cyberattack triggers a system-wide power loss. This may be extremely difficult and is mostly a manual operation for grid operators, since they must strike a fine balance among energy supply and demand; the entire system may shut down when there are too many clients demands or if power grows too rapidly. It is improbable that you will be able to recover a system-wide function known as "black start," but being able to do it safely is important to sustaining safe operation. For this purpose, the entire grid system can be intelligently supported with back-up batteries, as well as for cyber-attacks. It should be strengthened and harmonized with the digital twin concept. In this way, controllability is increased.

In addition, establishing working groups and creating a road map for the potential and future in this field will encourage private initiatives in the future and will give entrepreneurs the main materials necessary to get started.

REFERENCES

- Abdelsalam, A. K., Massoud, A. M., Ahmed, S., & Enjeti, P. N. (2011). High-performance adaptive Perturb and observe MPPT technique for photovoltaic-based microgrids. *IEEE Transactions on Power Electronics*, 26(4), 1010–1021. <https://doi.org/10.1109/TPEL.2011.2106221>
- Achieng, K. O. (2019). Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models. *Computers and Geosciences*, 133(September), 104320. <https://doi.org/10.1016/j.cageo.2019.104320>
- Akinsipe, O. C., Moya, D., & Kaparaju, P. (2021). Design and economic analysis of off-grid solar PV system in Jos-Nigeria. *Journal of Cleaner Production*, 287, 125055. <https://doi.org/10.1016/j.jclepro.2020.125055>
- Alagu, M., Ponnusamy, P., Pandarinathan, S., & Mohamed Ali, J. S. (2021). Performance improvement of solar PV power conversion system through low duty cycle DC-DC converter. *International Journal of Circuit Theory and Applications*, 49(2), 267–282. <https://doi.org/10.1002/cta.2918>
- Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian elastic net regression. *Communications in Statistics: Simulation and Computation*, 47(4), 1168–1178. <https://doi.org/10.1080/03610918.2017.1307399>
- Anderl, R., Haag, S., Schützer, K., & Zancul, E. (2021). Digital twin technology-An approach for Industrie 4.0 vertical and horizontal lifecycle integration. *IT - Information Technology*, 60(3), 125–132. <https://doi.org/10.1515/itit-2017-0038>
- Asimov, R. M. (2018). Digital twin in the analysis of a big data. *Big Data and ...*, (May). Retrieved from https://www.researchgate.net/profile/R_Asimov/publication/325038225_DIGITAL_TWIN_IN_THE_ANALYSIS_OF_A_BIG_DATA/links/5af2cc3e458515c283797f7a/DIGITAL-TWIN-IN-THE-ANALYSIS-OF-A-BIG-DATA.pdf
- Baboli, P. T., Babazadeh, D., & Kumara Bowatte, D. R. (2020). Measurement-based Modeling of Smart Grid Dynamics: A Digital Twin Approach. *2020 10th Smart Grid Conference, SGC 2020*, (Cim). <https://doi.org/10.1109/SGC52076.2020.9335750>
- Badal, F. R., Das, P., Sarker, S. K., & Das, S. K. (2019). A survey on control issues in renewable

- energy integration and microgrid. *Protection and Control of Modern Power Systems*, 4(1).
<https://doi.org/10.1186/s41601-019-0122-8>
- Bellia, H., Youcef, R., & Fatima, M. (2014). A detailed modeling of photovoltaic module using MATLAB. *NRIAG Journal of Astronomy and Geophysics*, 3(1), 53–61.
<https://doi.org/10.1016/j.nrjag.2014.04.001>
- Bhandari, B., Lee, K. T., Lee, C. S., Song, C. K., Maskey, R. K., & Ahn, S. H. (2014). A novel off-grid hybrid power system comprised of solar photovoltaic, wind, and hydro energy sources. *Applied Energy*, 133, 236–242. <https://doi.org/10.1016/j.apenergy.2014.07.033>
- Bosse, S., Maniry, D., Müller, K. R., Wiegand, T., & Samek, W. (2018). Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, 27(1), 206–219. <https://doi.org/10.1109/TIP.2017.2760518>
- bp. (n.d.). *Full report – Statistical Review of World Energy 2021*.
- bp Energy Outlook 2022. (n.d.).
- Del Moral, D. L., Barrado, A., Sanz, M., Lazaro, A., Fernandez, C., & Zumel, P. (2015). High efficiency DC-DC autotransformer forward-flyback converter for DMPPT architectures in solar plants. *Proceedings - 2015 9th International Conference on Compatibility and Power Electronics, CPE 2015*, 431–436. <https://doi.org/10.1109/CPE.2015.7231114>
- Du, J., Tu, Y., Dai, L. R., & Lee, C. H. (2016). A regression approach to single-channel speech separation via high-resolution deep neural networks. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 24(8), 1424–1437.
<https://doi.org/10.1109/TASLP.2016.2558822>
- Ebrahimi, A. (2019). Challenges of developing a digital twin model of renewable energy generators. *IEEE International Symposium on Industrial Electronics, 2019-June*, 1059–1066. <https://doi.org/10.1109/ISIE.2019.8781529>
- Faranda, R., & Leva, S. (2008). Energy comparison of MPPT techniques for PV Systems. *WSEAS Transactions on Power Systems*, 3(6), 446–455. Retrieved from <http://www.wseas.us/e-library/transactions/power/2008/27-545.pdf>
- Femia, N., Petrone, G., Spagnuolo, G., & Vitelli, M. (2004). Optimizing sampling rate of P&O MPPT technique. *PESC Record - IEEE Annual Power Electronics Specialists Conference*, 3, 1945–1949. <https://doi.org/10.1109/PESC.2004.1355415>
- Femia, Nicola, Petrone, G., Spagnuolo, G., & Vitelli, M. (2009). A technique for improving P&O MPPT performances of double-stage grid-connected photovoltaic systems. *IEEE Transactions on Industrial Electronics*, 56(11), 4473–4482.

- <https://doi.org/10.1109/TIE.2009.2029589>
- Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access*, 8, 108952–108971. <https://doi.org/10.1109/ACCESS.2020.2998358>
- G, D., & Singh, S. N. (2017). Selection of non-isolated DC-DC converters for solar photovoltaic system. *Renewable and Sustainable Energy Reviews*, 76(March), 1230–1247. <https://doi.org/10.1016/j.rser.2017.03.130>
- Gandoman, F. H., Ahmadi, A., Sharaf, A. M., Siano, P., Pou, J., Hredzak, B., & Agelidis, V. G. (2018). Review of FACTS technologies and applications for power quality in smart grids with renewable energy systems. *Renewable and Sustainable Energy Reviews*, 82(August 2016), 502–514. <https://doi.org/10.1016/j.rser.2017.09.062>
- Ghabeli Sani, S., Mohammadi, F., Banaei, M. R., & Farhadi-Kangarlu, M. (2019). Design and implementation of a new high step-up DC-DC converter for renewable applications. *International Journal of Circuit Theory and Applications*, 47(3), 464–482. <https://doi.org/10.1002/cta.2593>
- Haddad, K., Rahman, A., & Stedinger, J. R. (2012). Regional flood frequency analysis using Bayesian generalized least squares: A comparison between quantile and parameter regression techniques. *Hydrological Processes*, 26(7), 1008–1021. <https://doi.org/10.1002/hyp.8189>
- Hannan, M. A., Tan, S. Y., Al-Shetwi, A. Q., Jern, K. P., & Begum, R. A. (2020). Optimized controller for renewable energy sources integration into microgrid: Functions, constraints and suggestions. *Journal of Cleaner Production*, 256. <https://doi.org/10.1016/j.jclepro.2020.120419>
- Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. *Journal of the American Statistical Association*, 106(496), 1383–1393. <https://doi.org/10.1198/jasa.2011.tm09241>
- He, R., Chen, G., Dong, C., Sun, S., & Shen, X. (2019). Data-driven digital twin technology for optimized control in process systems. *ISA Transactions*, 95(xxxx), 221–234. <https://doi.org/10.1016/j.isatra.2019.05.011>
- He, T., Kong, R., Holmes, A. J., Nguyen, M., Sabuncu, M. R., Eickhoff, S. B., ... Yeo, B. T. T. (2020). Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*, 206, 116276. <https://doi.org/10.1016/j.neuroimage.2019.116276>

- Hernández-Callejo, L., Gallardo-Saavedra, S., & Alonso-Gómez, V. (2019). A review of photovoltaic systems: Design, operation and maintenance. *Solar Energy*, 188(March), 426–440. <https://doi.org/10.1016/j.solener.2019.06.017>
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251–257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
- International Renewable Energy Agency. (2021). Renewable Capacity Highlights. *Irena*, (March 2021), 1–3.
- IRENA. (2021). *Renewable Capacity Statistics*. *IRENA*.
- Ishwaran, H., & Lu, M. (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in Medicine*, 38(4), 558–582. <https://doi.org/10.1002/sim.7803>
- JRC Photovoltaic Geographical Information System (PVGIS) - European Commission. (2016, January 11). Retrieved from https://re.jrc.ec.europa.eu/pvg_tools/en/#PVP
- Karthikeyan, V., Rajasekar, S., Das, V., & Karuppanan, P. (2017). *20C96Af1377546664647906Dcc965D9E.Pdf*. <https://doi.org/10.1007/978-3-319-50197-0>
- Koç¹, A., Yağlı, H., Koç, Y., & Uğurlu, İ. (n.d.). *Dünyada ve Türkiye’de Enerji Görünümünün Genel Değerlendirilmesi*.
- Koop, G., & Poirier, D. J. (2004). Bayesian variants of some classical semiparametric regression techniques. *Journal of Econometrics*, 123(2), 259–282. <https://doi.org/10.1016/j.jeconom.2003.12.008>
- Kufeoğlu, S., Liu, G., Karim, A., & Pollitt, M. (2019). Digitalisation and New Business Models in Energy Sector. *EPRG Working Paper . Energy Policy Research Group, University of Cambridge*, (December), 1–39. Retrieved from <https://www.eprg.group.cam.ac.uk/wp-content/uploads/2019/06/1920-Text.pdf>
- Kumar, A., Chinnam, R. B., & Tseng, F. (2019). An HMM and polynomial regression based approach for remaining useful life and health state estimation of cutting tools. *Computers and Industrial Engineering*, 128, 1008–1014. <https://doi.org/10.1016/j.cie.2018.05.017>
- Lai, J. P., Chang, Y. M., Chen, C. H., & Pai, P. F. (2020). A survey of machine learning models in renewable energy predictions. *Applied Sciences (Switzerland)*, 10(17). <https://doi.org/10.3390/app10175975>
- Leela Nivashini, M., Madhumitha, K., Sindhu, S., & Maheswari, R. V. (2020). Design of high

- voltage transformer using finite element method. *International Journal of Scientific and Technology Research*, 9(3), 5351–5355.
- Linero, A. R. (2018). Bayesian Regression Trees for High-Dimensional Prediction and Variable Selection. *Journal of the American Statistical Association*, 113(522), 626–636. <https://doi.org/10.1080/01621459.2016.1264957>
- Liu, J., Liang, G., Siegmund, K. D., & Lewinger, J. P. (2018). Data integration by multi-tuning parameter elastic net regression. *BMC Bioinformatics*, 19(1), 1–9. <https://doi.org/10.1186/s12859-018-2401-1>
- Madni, A. M., Madni, C. C., & Lucero, S. D. (2019). Leveraging digital twin technology in model-based systems engineering. *Systems*, 7(1), 1–13. <https://doi.org/10.3390/systems7010007>
- Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big Data*, 4(1). <https://doi.org/10.1186/s40537-017-0077-4>
- Martín, L., Zarzalejo, L. F., Polo, J., Navarro, A., Marchante, R., & Cony, M. (2010). Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Solar Energy*, 84(10), 1772–1781. <https://doi.org/10.1016/j.solener.2010.07.002>
- Martínez, R., Bolea, Y., Grau, A., & Martínez, H. (2009). Fractional DC/DC converter in solar-powered electrical generation systems. *ETFA 2009 - 2009 IEEE Conference on Emerging Technologies and Factory Automation*, 1–6. <https://doi.org/10.1109/ETFA.2009.5347172>
- Mercadier, M., & Lardy, J. P. (2019). Credit spread approximation and improvement using random forest regression. *European Journal of Operational Research*, 277(1), 351–365. <https://doi.org/10.1016/j.ejor.2019.02.005>
- Mirzaee, A., & Moghani, J. S. (2020). Coupled Inductor-Based High Voltage Gain DC-DC Converter for Renewable Energy Applications. *IEEE Transactions on Power Electronics*, 35(7), 7045–7057. <https://doi.org/10.1109/TPEL.2019.2956098>
- Motlagh, N. H., Mohammadrezaei, M., Hunt, J., & Zakeri, B. (2020). Internet of things (IoT) and the energy sector. *Energies*, 13(2), 1–27. <https://doi.org/10.3390/en13020494>
- Muth, C., Oravecz, Z., & Gabry, J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology*, 14(2), 99–119. <https://doi.org/10.20982/tqmp.14.2.p099>
- Najafzadeh, M., Laucelli, D. B., & Zahiri, A. (2017). Application of model tree and

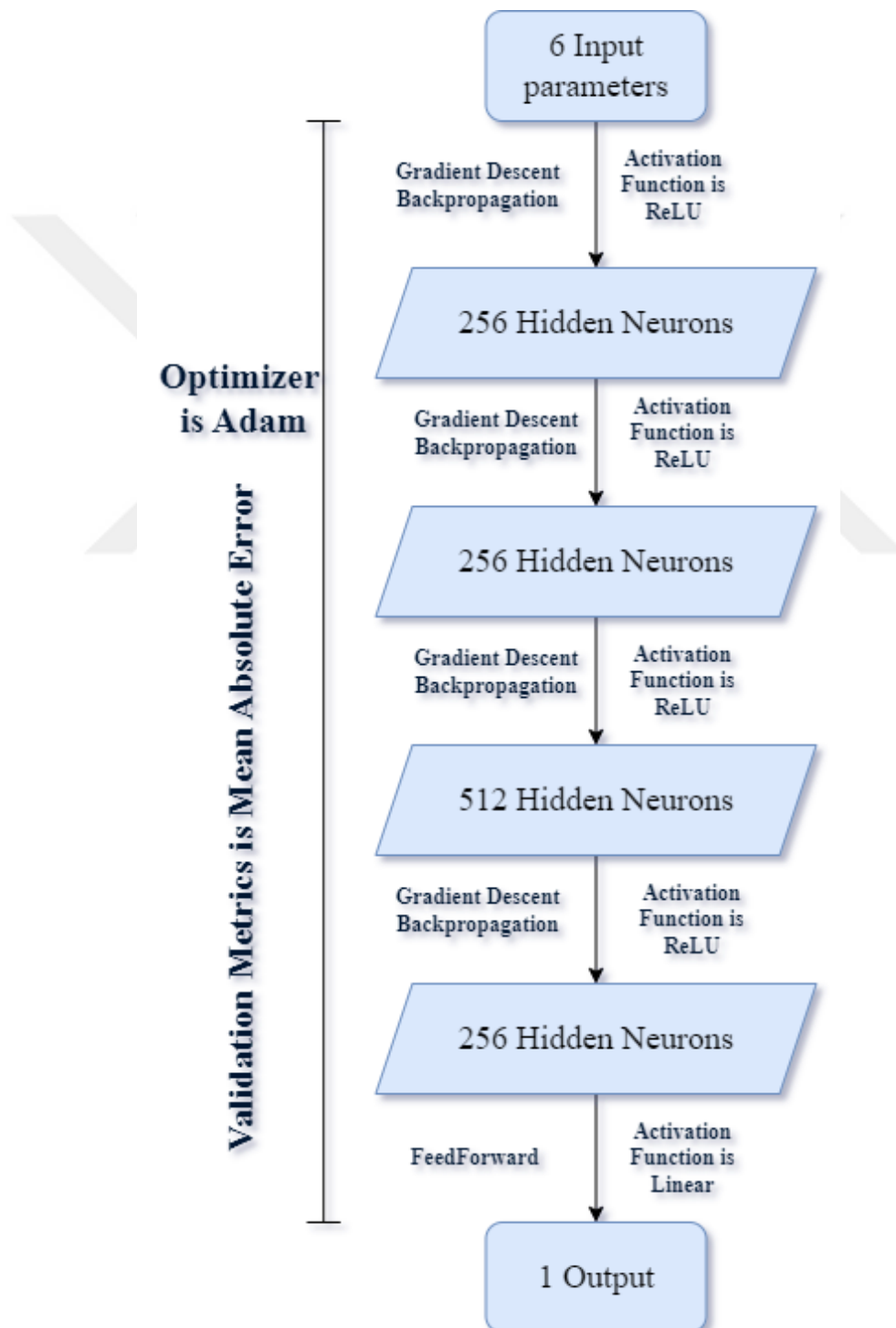
- Evolutionary Polynomial Regression for evaluation of sediment transport in pipes. *KSCE Journal of Civil Engineering*, 21(5), 1956–1963. <https://doi.org/10.1007/s12205-016-1784-7>
- Nguyen, V. H., Tran, Q. T., Besanger, Y., Jung, M., & Nguyen, T. L. (2022). Digital twin integrated power-hardware-in-the-loop for the assessment of distributed renewable energy resources. *Electrical Engineering*, 104(2), 377–388. <https://doi.org/10.1007/s00202-021-01246-0>
- Pan, Q., Qu, X., Liu, L., & Dias, D. (2020). A sequential sparse polynomial chaos expansion using Bayesian regression for geotechnical reliability estimations. *International Journal for Numerical and Analytical Methods in Geomechanics*, 44(6), 874–889. <https://doi.org/10.1002/nag.3044>
- Pimenta, F., Pacheco, J., Branco, C. M., Teixeira, C. M., & Magalhaes, F. (2020). Development of a digital twin of an onshore wind turbine using monitoring data. *Journal of Physics: Conference Series*, 1618(2). <https://doi.org/10.1088/1742-6596/1618/2/022065>
- Pollack, J. B. (1989). Perceptrons: An Introduction to Computational Geometry, Expanded Edition. *J. Math. Psychology*, 33(1969), 358–365.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31(4), 437–448. <https://doi.org/10.3102/10769986031004437>
- Rahman, M. M., Shakeri, M., Tiong, S. K., Khatun, F., Amin, N., Pasupuleti, J., & Hasan, M. K. (2021). Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability (Switzerland)*, 13(4), 1–28. <https://doi.org/10.3390/su13042393>
- Rezaie-Balf, M., & Kisi, O. (2018). New formulation for forecasting streamflow: Evolutionary polynomial regression vs. extreme learning machine. *Hydrology Research*, 49(3), 939–953. <https://doi.org/10.2166/nh.2017.283>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Salcedo-Sanz, S., Cornejo-Bueno, L., Prieto, L., Paredes, D., & García-Herrera, R. (2018). Feature selection in machine learning prediction systems for renewable energy applications. *Renewable and Sustainable Energy Reviews*, 90(March), 728–741. <https://doi.org/10.1016/j.rser.2018.04.008>

- Selmi, T., & Belghouthi, R. (2017). A novel widespread Matlab/Simulink based modeling of InGaN double hetero-junction p-i-n solar cell. *International Journal of Energy and Environmental Engineering*, 8(4), 273–281. <https://doi.org/10.1007/s40095-017-0243-7>
- Sera, D., Mathe, L., Kerekes, T., Spataru, S. V., & Teodorescu, R. (2013). On the perturb-and-observe and incremental conductance mppt methods for PV systems. *IEEE Journal of Photovoltaics*, 3(3), 1070–1078. <https://doi.org/10.1109/JPHOTOV.2013.2261118>
- Shuvho, M. B. A., Chowdhury, M. A., Ahmed, S., & Kashem, M. A. (2019). Prediction of solar irradiation and performance evaluation of grid connected solar 80KWp PV plant in Bangladesh. *Energy Reports*, 5, 714–722. <https://doi.org/10.1016/j.egyr.2019.06.011>
- Solar Power Europe. (2017). Digitalisation & Solar Task Force Report. *Solar Power Europe*, 40. Retrieved from https://www.solarpowereurope.org/wp-content/uploads/2018/09/Digitalisation_and_Solar_report_SolarPower_Europe_MEDIUM_RES.pdf
- Testad, I., Corbett, A., & Aarsland, D. et al. (2013). ORE Open Research Exeter. *Journal of Cleaner Production*, 0–48.
- Thomas, S., & Nithiyananthan, K. (2018). A novel method to implement MPPT algorithms for PV panels on a MATLAB/ SIMULINK environment. *Journal of Advanced Research in Dynamical and Control Systems*, 10(4), 31–40.
- Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 106, 234–240. <https://doi.org/10.1016/j.sbspro.2013.12.027>
- van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, 89, 31–50. <https://doi.org/10.1016/j.jmp.2018.12.004>
- Vimalarani, C., & Kamaraj, N. (2015). Modeling and performance analysis of the solar photovoltaic cell model using Embedded MATLAB. *Simulation*, 91(3), 217–232. <https://doi.org/10.1177/0037549714568535>
- Xu, X., Wei, Z., Ji, Q., Wang, C., & Gao, G. (2019). Global renewable energy development: Influencing factors, trend predictions and countermeasures. *Resources Policy*, 63(April). <https://doi.org/10.1016/j.resourpol.2019.101470>
- Xu, Y., Du, J., Dai, L. R., & Lee, C. H. (2015). A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 23(1), 7–19. <https://doi.org/10.1109/TASLP.2014.2364452>

- Yuchi, W., Gombojav, E., Boldbaatar, B., Galsuren, J., Enkhmaa, S., Beejin, B., ... Allen, R. W. (2019). Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. *Environmental Pollution*, 245, 746–753. <https://doi.org/10.1016/j.envpol.2018.11.034>
- Zainal, N. A., Ajisman, & Yusoff, A. R. (2016). Modelling of Photovoltaic Module Using Matlab Simulink. *IOP Conference Series: Materials Science and Engineering*, 114(1). <https://doi.org/10.1088/1757-899X/114/1/012137>
- Zhang, J., Ma, G., Huang, Y., sun, J., Aslani, F., & Nener, B. (2019). Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Construction and Building Materials*, 210, 713–719. <https://doi.org/10.1016/j.conbuildmat.2019.03.189>
- Zhang, Z., Lai, Z., Xu, Y., Shao, L., Wu, J., & Xie, G. Sen. (2017). Discriminative Elastic-Net Regularized Linear Regression. *IEEE Transactions on Image Processing*, 26(3), 1466–1481. <https://doi.org/10.1109/TIP.2017.2651396>
- Zheng, J., Zhang, H., Dai, Y., Wang, B., Zheng, T., Liao, Q., ... Song, X. (2020). Time series prediction for output of multi-region solar power plants. *Applied Energy*, 257(October 2019), 114001. <https://doi.org/10.1016/j.apenergy.2019.114001>
- Zheng, X., Jiang, Z., Ying, Z., Song, J., Chen, W., & Wang, B. (2020). Role of feedstock properties and hydrothermal carbonization conditions on fuel properties of sewage sludge-derived hydrochar using multiple linear regression technique. *Fuel*, 271(February), 117609. <https://doi.org/10.1016/j.fuel.2020.117609>
- Zheng, Y., Yang, S., & Cheng, H. (2019). An application framework of digital twin and its case study. *Journal of Ambient Intelligence and Humanized Computing*, 10(3), 1141–1153. <https://doi.org/10.1007/s12652-018-0911-3>
- Zou, H., & Hastie, T. (2003). Regression Shrinkage and Selection via the Elastic Net, with Applications to Microarrays. *Journal of the Royal Statistical Society: Series B*, 67(1), 301–320. Retrieved from <http://webdocs.cs.ualberta.ca/~mahdavif/ReadingGroup/Papers/10.1.1.9.6188.pdf>

ANNEXES

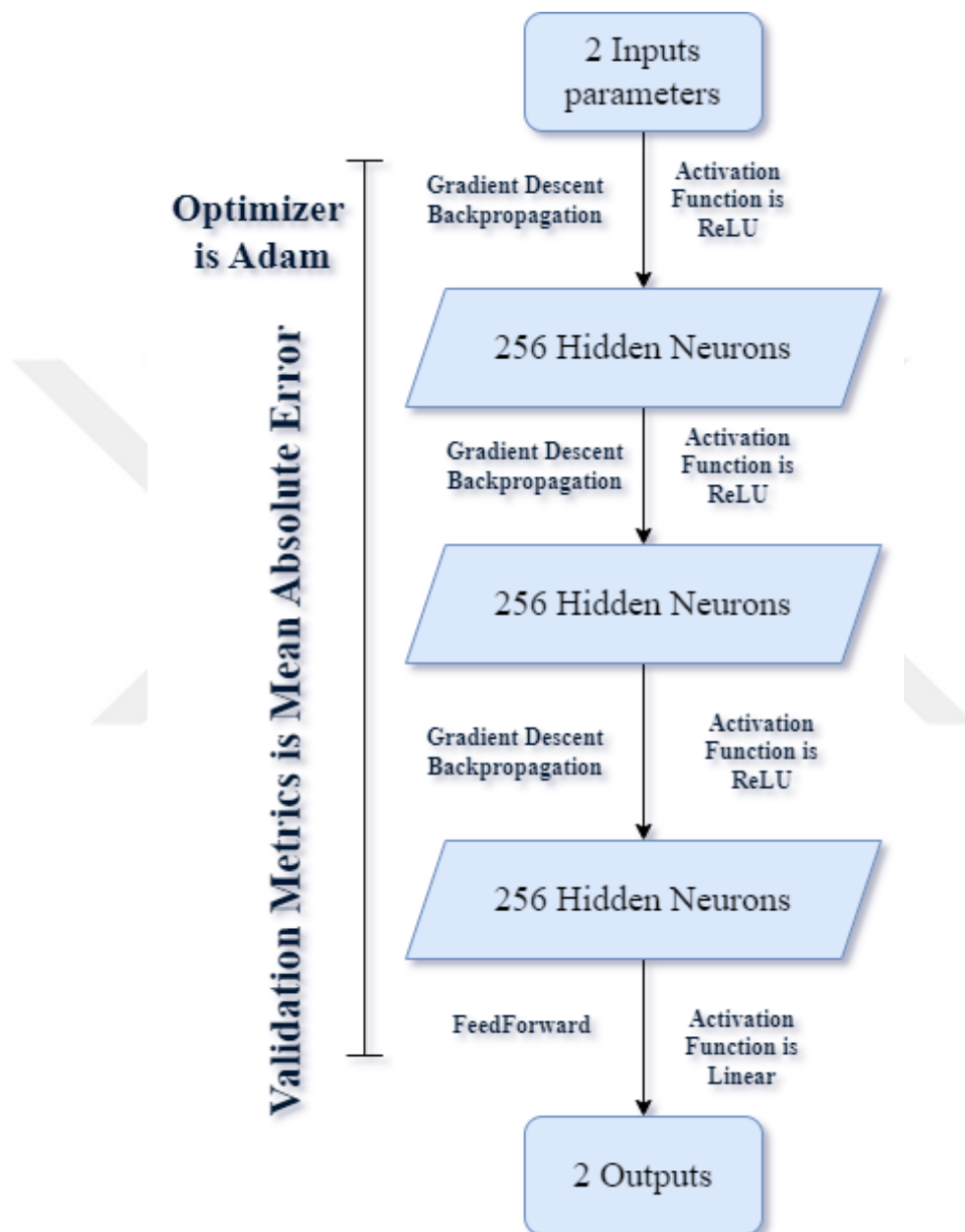
Annex-A: The Architectural Structure Of The ANN Used In The PV Panel Part



Annex-B: Comparing And Examining The Prediction Results Of The PV Panel Part

Number	Real Power Value of PV (W)	Prediction From DNN (W)	Error Values in Watts(W)	Error Rate (%)
1	0.00	-0.85	0.85	42085.39
2	124541.86	124686.16	144.29	0.12
3	-0.04	-0.13	0.09	219.08
4	-0.33	-0.23	0.10	30.27
5	-1.12	-0.37	0.74	66.59
6	-0.33	-0.17	0.16	48.34
7	50562.11	50546.37	15.74	0.03
8	0.00	-0.18	0.18	5188.46
9	114215.89	114405.47	189.85	0.17
10	-0.07	-0.15	0.08	113.92
11	4554.72	4558.69	3.97	0.09
12	4317.16	4323.64	6.48	0.15
13	-0.67	-0.14	0.53	79.75
14	-4.28	-0.20	4.08	95.34
15	80747.03	80856.23	109.20	0.14
16	5273.52	5263.14	10.37	0.20
17	-3.05	-0.27	2.78	91.20
18	119260.59	11943.81	174.22	0.15
19	8093.78	8101.59	7.80	0.10
20	29013.17	29087.39	74.22	0.26
21	93281.15	93356.89	75.74	0.08
22	-0.06	-0.18	0.13	224.97
23	-8.81	-0.21	8.60	97.62
24	16940.47	16907.39	33.07	0.20
25	114451.09	114482.92	31.84	0.03

Annex-C: The Architectural Structure Of The ANN Used In The DC-DC Converter Part



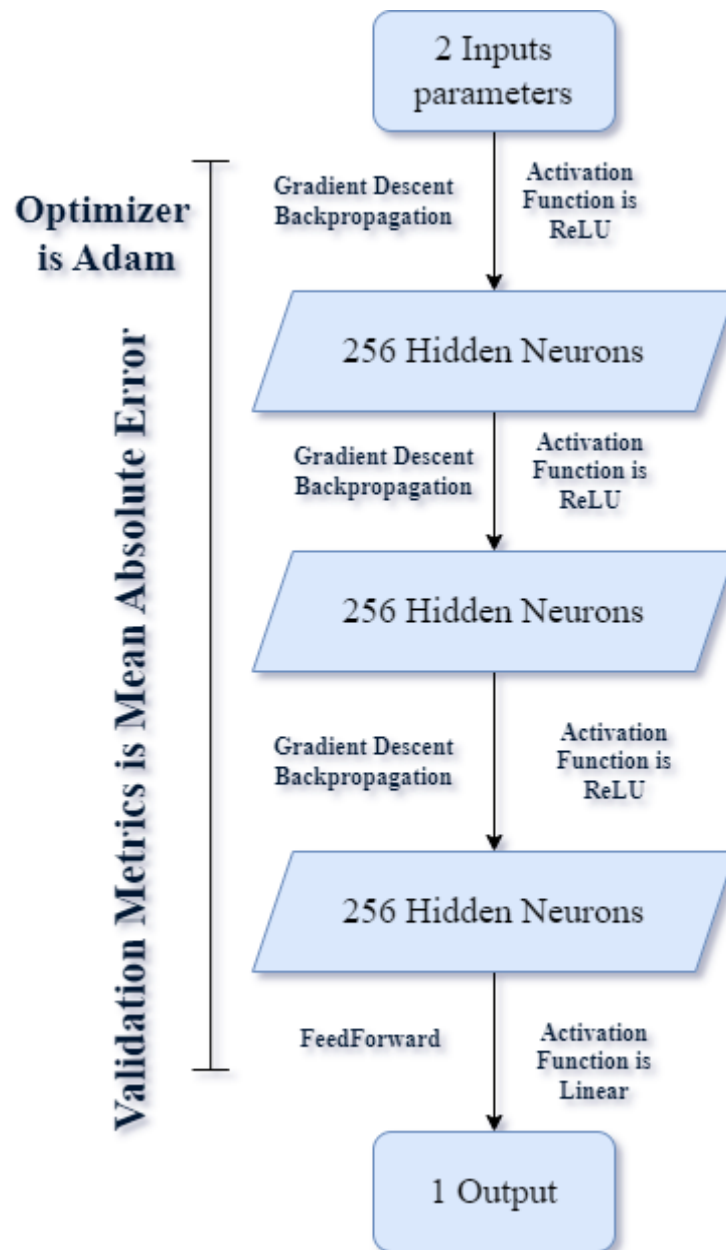
**Annex-D: Comparing And Examining The Prediction Results Of The DC-DC Converter
Part**

Number	Real Voltage Value of DC (I)	Prediction of Voltage Value (I)	Error Values in Voltage (I)	Real Current Value of DC (V)	Prediction of Current Value (V)	Error Values in Current (V)
1	-0.13	-0.58	0.45	499.02	497.98	1.04
2	0.47	-0.59	1.06	499.02	497.96	1.06
3	1.10	-0.60	1.69	499.04	497.94	1.10
4	1.71	-0.61	2.33	499.06	497.88	1.18
5	2.28	-0.63	2.91	499.08	497.82	1.26
6	2.62	-0.65	3.26	499.11	497.77	1.34
7	3.06	-0.67	3.73	499.13	497.74	1.39
8	3.22	-0.70	3.92	499.15	497.74	1.40
9	73.17	73.64	0.48	500.00	500.79	0.80
10	137.15	137.70	0.55	500.00	500.67	0.67
11	184.00	184.97	0.98	500.00	501.10	1.10
12	210.84	212.15	1.32	500.00	502.06	2.07
13	215.89	217.23	1.34	500.00	502.26	2.27
14	202.47	203.67	1.20	500.00	501.87	1.88
15	167.64	168.23	0.59	500.00	500.97	0.97
16	113.79	114.29	0.50	500.00	500.60	0.60
17	17.40	17.86	0.45	500.00	499.45	0.55
18	-5.44	-3.75	1.69	500.41	499.24	1.17
19	-6.38	-4.77	1.61	500.06	498.84	1.22
20	-6.12	-3.87	2.25	499.70	498.32	1.38
21	-2.97	-0.83	2.14	499.17	497.76	1.41
22	-0.60	-0.64	0.03	499.07	497.82	1.25
23	1.34	-0.64	1.99	499.09	497.79	1.30
24	2.92	-0.73	3.65	499.17	497.75	1.42

25	4.22	-1.19	5.41	499.29	497.79	1.50
----	------	-------	------	--------	--------	------



Annex-B: The Architectural Structure Of The ANN Used In The PV Grid Part



Annex-F: Comparing And Examining The Prediction Results Of The Grid Part

Number	Real Power Value of Grid (KW)	Prediction From DNN (KW)	Error Values in Kilowatts(KW)	Error Rate (%)
1	-0.29	-0.31	0.01	4.31
2	36.42	36.44	0.01	0.05
3	68.34	68.41	0.07	0.10
4	90.37	90.54	0.16	0.18
5	103.34	103.57	0.23	0.22
6	106.66	106.91	0.24	0.23
7	99.62	99.83	0.25	0.21
8	81.40	81.52	0.12	0.15
9	43.86	43.89	0.02	0.04
10	4.48	4.55	0.07	1.65
11	-0.33	-0.33	0.00	0.15
12	-0.31	-0.33	0.01	3.77
13	-0.30	-0.32	0.02	5.14
14	-0.30	-0.31	0.01	4.04
15	-0.29	-0.30	0.01	4.04
16	-0.29	-0.30	0.01	4.16
17	-0.28	-0.29	0.01	4.29
18	-0.28	-0.29	0.01	4.36
19	-0.28	-0.29	0.01	3.68
20	-0.27	-0.29	0.02	4.11
21	-0.27	-0.29	0.02	4.45
22	-0.27	-0.29	0.02	4.82
23	-0.27	-0.28	0.01	5.27
24	-0.27	-0.28	0.01	5.81
25	-0.27	-0.28	0.01	6.19

Annex-G: ML Regression Models, The Parameters And Values Used To Train Them

Type of Regression	Which part of the system	Number of inputs and outputs	Parameters used to build the model	Optimizer used to initialize the model	Number of the iterations	Error type used to train the model
DNN Regression	PV Panel Part	6 inputs, 256,256,512,256 Hidden, 1 Output	Gradient Descent - Backpropagation The activation function is ReLU or Linear	The optimizer of the models is Adam	500 Iterations	MSE
	DC-DC Converter Part	2 inputs, 256,256, 256 Hidden, 2 Outputs				
	Grid Part	2 inputs, 256,256, 256 Hidden, 1 Output				
Random Forest Regression	PV Panel Part	6 inputs and 1 output	The estimator* is between 100 to 1200 Maximum depth* is between 5 to 30 Samples split* = [2,5,10,15,100] Samples Leaf* = [1,2,5,10]	Feature = Auto or Sqrt	100 Iterations	MSE
	DC-DC Converter Part	2 inputs and 2 Outputs				
	Grid Part	2 Inputs and 1 Output				

ElasticNet Regression	PV Panel Part	6 inputs and 1 output	Alpha** = [0.0001,0.001,0.01,0.1,0.3, 0.5,0.7,1]	Cross Validation is 10	100 Iterations	MSE
	Grid Part	2 Inputs and 1 Output	L1 Ratio** = [0.2,0.4,0.6,0.8] Lambda** is 1e-3			
Bayesian Regression	PV Panel Part	6 inputs and 1 output	Lambda is 4 Alpha is 0.3 Beta is 1	-	100 Iterations	MSE
	Grid Part	2 Inputs and 1 Output				
3 rd Degree Polynomial Regression	PV Panel Part	6 inputs and 1 output	The coefficient of determination is 0.999 The intercept is 1.005	-	100 Iterations	MSE
	Grid Part	2 Inputs and 1 Output				
Linear Regression	PV Panel Part	6 inputs and 1 output	The coefficient of determination is 0.92 The intercept is 1.20	-	100 Iterations	MSE
	Grid Part	2 Inputs and 1 Output				

*Estimator is the number of trees in a random forest

*Feature is the number of features to consider at every split

*Maximum depth is the maximum number of levels in the tree

*Samples split is the minimum number of samples required to split a node

*Samples Leaf is the minimum number of samples required at each leaf node

** L1 Ratio is called a mixing parameter

** Alpha is specified to specify how many weights are given from each of the L1 and L2 penalties

**Lambda is provided to alter the weighting of the sum of both penalties in relation to the loss function

s