Implementation of Machine Learning Models for Predicting Hydrogen Production from Renewable Energy

MS Thesis
by
Adnan Ayub

CIIT/FA23-REE-002/LHR

COMSATS University
Islamabad Pakistan

Fall 2023

Implementation of Machine Learning Models for Predicting Hydrogen Production from Renewable Energy

A thesis submitted to
COMSATS University Islamabad

In partial fulfillment
of the requirement for the degree of

Master of Science
in
Electrical Engineering

by
Adnan Ayub

CIIT/FA23-REE-002/LHR

Department of Electrical Engineering
Faculty of Science

COMSATS University Islamabad
Pakistan

Fall 2023

Thesis title

This thesis is submitted to the department of Electrical Engineering in partial fulfillment of the requirement for the award of degree of Master of Science in Electrical Engineering.

Name	Registration Number
Adnan Ayub	CIIT/FA23-REE-002/LHR

Supervisor

Dr. Muhammad Yaqoob Javed

Associate Professor, Department of Electrical Engineering

COMSATS University Islamabad, Lahore Campus

Dec, 2025

Certificate of Approval

This thesis is titled

Implementation of Machine Learning Models for Predicting
Hydrogen Production from Renewable Energy

Adnan Ayub

CIIT/FA23-REE-002/LHR

Has been approved

For COMSATS University Islamabad, Lahore Campus.

External Examiner:____

Dr Yasir Saleem
University of Engineering and
Technology (UET), Lahore

Supervisor:____

Dr. Muhammad Yaqoob Javed
Department of Electrical Engineering, (CUI) Lahore Campus

Head of Department:____

Dr. Mirza Tariq Humayun
Department of Electrical Engineering, (CUI) Lahore Campus

Author’s Declaration

I, Adnan Ayub, CIIT/FA23-REE-002/LHR, hereby declare that I have produced the work presented in this thesis during the scheduled period of study. I also declare that I have not taken any material from any source except referred to wherever due to that amount of plagiarism is within an acceptable range. If a violation of HEC rules on research has occurred in this thesis, I shall be liable to punishable action under the plagiarism rules of HEC.

Date:______________ ____________________________

Adnan Ayub

CIIT/FA23-REE-002/LHR

Certificate

It is certified that Adnan Ayub, CIIT/FA23-REE-002/LHR, has carried out all the work related to this thesis under my supervision at the Department of Electrical Engineering, COMSATS University Islamabad, Lahore Campus, and the work fulfills the requirements for the award of the MS degree.

Date:______________

Supervisor

____

Dr. Muhammad Yaqoob Javed,
Associate Professor,
Electrical Engineering,
COMSATS University Islamabad
Lahore Campus

Dedication

This research report is dedicated to my parents, whose constant prayers, sacrifices, and encouragement have been the foundation of my academic journey. I am deeply grateful to my teachers and mentors for their guidance, knowledge, and support throughout this research work. I also dedicate this report to my friends and colleagues who provided motivation and assistance during challenging times. Above all, I thank Almighty Allah for granting me the strength, patience, and wisdom to complete this research successfully.

Acknowledgements

Praise be to ALLAH, the Cherisher and Lord
of the World, Most gracious and Most Merciful.

First and foremost, I would like to thank ALLAH Almighty (the most beneficent and most merciful) for giving me the strength, knowledge, ability, and opportunity to undertake this research study and to preserve and complete it satisfactorily. Without the countless blessings of ALLAH Almighty, this achievement would not have been possible. May His peace and blessings be upon His messenger, Hazrat Muhammad (PBUH), upon his family, companions, and whoever follows him. My heartfelt gratitude to Hazrat Muhammad (PBUH), who is forever a source of guidance and knowledge for humanity as a whole. In my journey towards this degree, I have found a teacher, an inspiration, a role model, and a pillar of support in my life, my kind.

Adnan Ayub
CIIT/FA23-REE-002/LHR

Abstract

The world continues to be largely reliant on fossil fuels, such as coal, oil and natural gas, which are major contributors of greenhouse gases, air pollution and climate change caused by the release of gases like CO_2, NO o and SO_2. Even though renewable energy sources such as solar and wind provide a cleaner substitute, their nature as intermittent and weather-dependent sources become a big problem in terms of large scale and long-term energy storage making traditional battery systems economically impractical. Consequently, the utilization of the surplus renewable energy through the process of water electrolysis to create green hydrogen has become a viable and alternative way of storing energy in the long term.

This work suggests a two-stage machine learning-based predictive model of hydrogen production when using renewable energy, based on real-world working data of a 40.5 MW grid-connected photovoltaic (PV) power facility. The initial step involves predicting photovoltaic power output from meteorological variables using multiple regression and deep learning models, such as Support Vector Regression (SVR), Random Forest, Decision Tree, and Gated Recurrent Units (GRU). The second stage is to incorporate the predicted solar energy into a hybrid electrochemical model of hydrogen production, and the same machine learning models serve as data-driven correction models to more effectively predict hydrogen yield. It is a two-stage method that integrates both physical modelling and machine learning to model nonlinear system behaviour and real-world losses of operation. The findings show that SVR was always better in both phases than the other models with an R 2 value of 0.965 in photovoltaic power prediction and 0.968 in hydrogen production prediction. The given framework minimized the mistake in the production of hydrogen annually to about 3 percent, which is much better than theoretical models and deep learning alternatives.

Moreover, a Sobol based global sensitivity analysis revealed that Global Horizontal Irradiance (GHI) made the greatest contribution to the uncertainty in the hydrogen production process, then AC power output and temperature at the module. The outcomes of these studies support the premise that the two stages proposed framework is a viable, precise, and large-scale solution to real-time forecasting, optimization of the system, and successful grid integration of green hydrogen system.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Modern lives depend on electricity, and our homes and industries rely on it. Our primary sources of electricity are fossil fuels (coal, oil, and gas), nuclear, solar, and wind. But fossil fuels, coal, oil, and gas are the dominant sources due to their high density, and infrastructure is developed. Still, they emit harmful gases such as CO_2, SO_2, and NOX, contributing to air pollution. Fossil fuels are limited and environmentally destructive, driving a global shift toward renewable energy sources like solar, wind, hydro, and geothermal. These green alternatives produce no air pollution. They are widely available, reduce dependence on foreign energy, and create jobs in the renewable energy sector. Countries are investing in these sources and setting pollution reduction goals for a sustainable future [1]. As fossil fuels are environmentally harmful, the world is now moving toward clean, green energy sources such as solar (photovoltaic cells), wind, hydro, and geothermal. All these energies are clean and green, emitting no harmful gases. It helps countries reduce their dependence on foreign energy and creates jobs in the local industry. However, renewable energies are not constant and are not available all the time; for example, solar energy is only available during the day, and wind energy is not available all the time [2]. To solve this problem, a storage system is needed. Lithium-ion batteries are, so far, the best batteries for storing energy, but at that level, they cannot provide backup for the whole city and are not cost-effective [3]. One solution is hydrogen energy. Hydrogen can serve as an energy carrier, storing excess electricity from renewable energy sources and releasing it as needed.

Some of the incredible sources of renewable energy, like solar and wind, exhibit high temporal variability not only in terms of intermittency but also in daily generation patterns. Usually, solar photovoltaic systems produce the most power during the daytime hours, while wind energy generation is greatest in the evening and at night. The supply of renewable energy often exceeds actual demand during peak production periods leading to curtailment or wastage due to grid constraints and the lack of large scale storage options. This mismatch between generation and consumption of power can be explained by one of the biggest inefficiencies in introducing renewable power and a significant barrier to high rates of renewable power introduction into the modern power systems [4].

The way out of this would also be possible in the process of electrolysing water to create green hydrogen, so that they can turn the excess renewable energy into a form of energy carrier, the green hydrogen, which is easily transportable and does not emit any carbon, which can be stored. The surplus solar power in the day and the surplus wind power in the night can channeled to the electrolyzers to generate other hydrogen as opposed to wastage. This hydrogen can be stored for long periods. It could be used in the future as a clean source of transportation, in industrial applications (such as steel and chemical production), or converted back to electricity using fuel cells when renewable energy production is minimal. Hydrogen not only increases the overall efficiency of the system and supports grid stability, deep decarbonization, and sector coupling by serving as a temporal energy bridge between variable renewable energy supply and uninterrupted energy demand. Therefore, hydrogen production from renewable energy sources can turn the limitation of intermittency into an opportunity and create a resilient ecosystem, as explained in Figure 1-1, which represents a sustainable energy ecosystem.

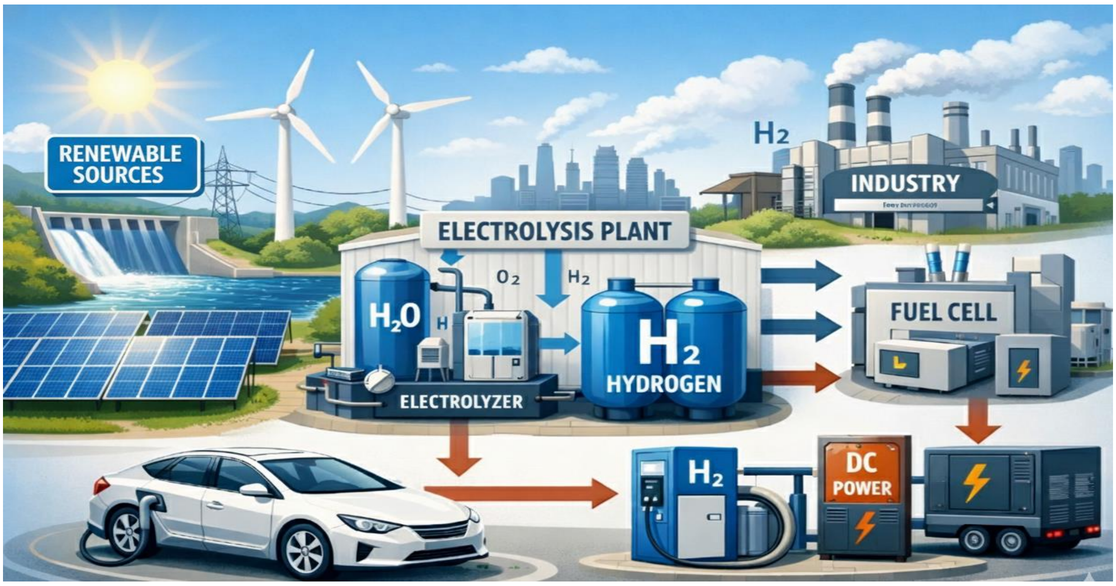

Figure 1- 1Hydrogen Production from Renewable Energy [5]

This is done with electrolysis. Electricity is used to divide water (H_2O) into hydrogen (H_2) and oxygen (O_2) by the process of electrolysis [6]. When renewable energy is used in the method shown in Figure 1, the produced hydrogen is called "green hydrogen". This hydrogen stored transported and utilized in fuel cells to produce electricity once more or utilized by industry itself. This is what makes hydrogen a clean and versatile energy carrier that can be used to decarbonize most industries. Electrolysis is an efficient and easy process. During the electrolysis process, a current of electricity is passed through the water by use of two electrodes; an anode and a cathode.Water is oxidized to oxygen gas and hydrogen ions (H^+). Hydrogen ions at the cathode accept electrons, resulting in the formation of hydrogen gas. There are different types of electrolyzers based on the electrolyte used: alkaline, proton exchange membrane (PEM), and solid oxide electrolyzers (SOE). The electrolysis process is chemically mentioned below.

$$2H_2O \rightarrow 2H_2 + O_2 \tag{1}$$

$$2H_2O(l) \overset{\text{electric current}}{\rightarrow} 2H_2(g) + O_2(g) \tag{2}$$

Ions in solution: When water is slightly ionized (or with an electrolyte added like H_2SO_4 for better conductivity):

$$H_2O \leftrightarrow H^+ + OH^- \tag{3}$$

At the electrodes:

1. Cathode (negative electrode, reduction occurs):

$$2H^{+} + 2e^{-} \rightarrow H_2(g) \tag{4}$$

2. Anode (positive electrode, oxidation occurs):

$$4OH^{-} \rightarrow O_2(g) + 2H_2O + 4e^{-}z \tag{5}$$

Now, there is a problem with generating power for the electrolysis process. One of the environmentally friendly and cost-effective methods is to power the electrolysis process with photovoltaic cells, producing electricity to meet power needs. To initiate this process, a photovoltaic (PV) system is required, which converts sunlight into direct current (DC) electricity. PVs are made up of semiconductors, mostly of silicon. When sunlight or solar radiation strikes the surface of the PV module, the silicon cells are excited, and electrons begin to move, producing direct current [7]. This makes direct current, which can then be supplied directly to the electrolyzer for electrolysis. An electrolyzer uses direct current to split the molecules of water (H2O) into separate molecules of hydrogen (H2 and oxygen (O2). Figure 1-2 shows a simple diagram of the electrolysis process, highlighting the flow of current and hydrogen production.

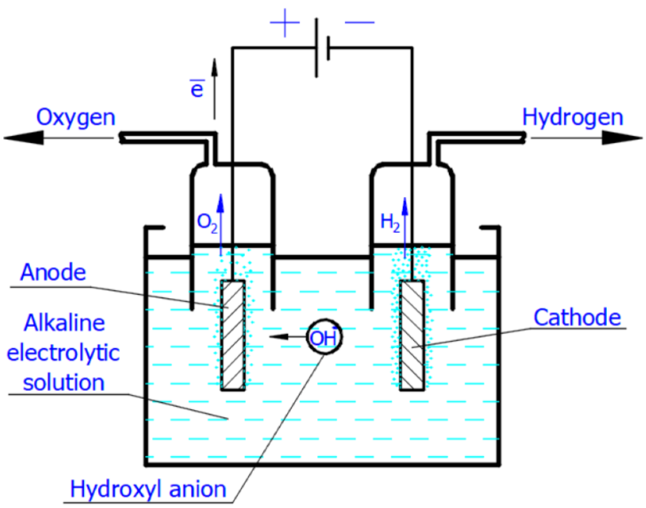

Figure 1- 2The process of Electrolysis [8]

Although renewable energy-based hydrogen production is technically feasible, the highly fluctuating and unpredictable solar irradiance and renewable energy input pose a significant challenge when planning, operating, and optimizing electrolysis systems. It can therefore argued that predicting hydrogen production is one of the essential steps to efficient scheduling of the systems, energy management, and cost-effective functioning. The nonlinear interactions of weather conditions, power generation, and hydrogen yield are more likely to overlooked in unconventional analytical and physics based models, and the data-driven predictive models are required. This study aims at predicting hydrogen production using machine-learning algorithms on a renewable energy source such as solar energy. Conventional techniques will probably fail to represent the non-linear and complicated interactions that are involved in the electrolysis process [9]. The ability of machine learning to process large amounts of data and find the latent patterns forms an attainable solution to developing accurate predictive models. In the study, the machine learning algorithms will train on the specific features such as energy inputs, weather, system parameters, and others and predict the output of hydrogen. The results of the models will evaluate with the conventional evaluation metrics including R2 Score, Mean Squared Error (MSE), and Root Mean Square Error (RMSE) to identify the optimal outcome at the least cost. The main aim of the research is to enhance the accuracy and reliability of the predictions of hydrogen production. This will facilitate greater planning, simplified management of their systems and smoother integration of hydrogen into the current renewable energy systems [10].

Research gap:

The machine learning method has widely been utilized in the literature to predict the generation of renewable energy, including solar, wind power, and to model the production of hydrogen as an independent process. These two areas are however to a large extent studied separately. The mechanism to directly link the inputs of variable renewable energy and the consequent hydrogen generation has not been adequately considered through the integrated machine learning methods. Specifically, the studies that apply and compare several machine learning models to predict the output of hydrogen using the availability of renewable energy and working conditions are insufficient. This drawback prevents the realistic modeling of real-life hydrogen systems that are powered by renewable energy. Thus, a predictive framework should be provided that includes machine learning to solve the problem of the disconnect between renewable energy generation and hydrogen production.

Research objectives:

The objectives of this research work are as follows:

- For this research we will find a dataset of solar energy production and electrolysis process, then data will be prepared for machine learning, and selecting suitable input and output parameters.

- Implementation of machine learning algorithms for prediction of hydrogen production. 3. Tuning machine learning algorithms' hyperparameters to get optimized results.

- Comparisons of results based on R2 Score, Mean Squared Error (MSE), and Root Mean Square Error (RMSE), and selection of the most accurate algorithm to predict hydrogen production from renewable energy.

Problem statement:

Hydrogen generation from alternative energy sources such as wind, sun, and Ocean Thermal Energy Conversion (OTEC) is a green process in contrast to fossil fuels. It is challenging to forecast hydrogen generation due to renewable sources being dynamic in nature. Conventional modeling software like Aspen Plus and EES cannot cope with dynamic, nonlinear conditions and big data. Machine learning offers a promising solution through predictive analytics. However, studies assessing which ML models are best suited for hydrogen production forecasting have been explored to a limited extent. This research aims to use machine learning for the forecasting of hydrogen production.

Chapter 2
Literature

2.1. Introduction

The growing power requirement around the world and the environmental impact of fossil fuel consumption in power generation has been the key factors behind the transition to renewable and sustainable sources of energy. The bulk of the carbon dioxide and other greenhouse gases emitted all over the world are a product of combustion of so-called conventional sources of energy especially coal, oil and natural gas. They are the main sources of electricity in the world due to their high energy density and most of the countries already possess them, but they are the largest origins of greenhouse gasses and air pollution. This fluctuation is an issue to massive energy storage and power grid stability. Hydrogen has also suggested as a viable energy carrier to deal with the storage and flexibility issues of the renewable energy systems [11]. Green hydrogen, which is a product made out of renewable electricity through water electrolysis, offers zero-emission energy storage, transportation, and end use in the power sector, industrial sector, and transportation industry. Recent literature has aimed at improving the efficiency and reliability, and economic feasibility of green hydrogen systems as well as through the use of advanced computational modelling and machine learning (ML) algorithms. The given literature review is particularly matched to the photovoltaic (PV)-powered hydrogen generation systems and ML-powered hydrogen prediction models, which will be at the core of the suggested two-stage predictive approach.

2.2 Renewable Energy to Green Hydrogen Production.

Electrolytic hydrogen generation has been identified as one of the key ways of decarbonizing the energy industry using renewable electricity as the electricity source [12]. The primary method of electrolyzing water into green hydrogen entails the use of electrical energy to split the water molecules into hydrogen and oxygen. Using renewable energy as the power source would lead to zero direct carbon emissions. Electrochemical reactions occur between two electrodes, one an anode and the other a cathode, and the process's efficiency depends heavily on operating conditions, the type of electrolyte used, and the stability of the power supply. Various electrolysis technologies can be applied to renewable integration. Alkaline electrolyzers are already technologically advanced and cost-efficient but have a low dynamic response. The most important feature of the present research is Proton Exchange Membrane (PEM) electrolyzers that have the ability to respond quickly with high current density and compatibility with fluctuating renewable power sources at a high cost of capital due to the use of the noble-metal catalysts [13]. SOEs are very efficient at high temperatures but there are durability problems in dynamic operation.

Hydrogen production systems with photovoltaics amplify the solar irradiance into electrical energy, which is directly fed to electrolyzers. The lack of grid electricity in PV E systems enables hydrogen production on a decentralized basis. But PV performance is susceptible to meteorological conditions, including global horizontal irradiance (GHI), module temperature, and ambient conditions, leading to nonlinear, time-dependent hydrogen production. Several studies have shown that PV-E systems are viable for hydrogen production in areas with high solar potential [14]. The system's performance analysis shows that proper modeling of PV performance and electrolyzer efficiency is required for reliable hydrogen predictions and economic optimization**.**

2.2.1 **Thermodynamic and Electrochemical Modeling.**

Thermodynamic modeling determines the theoretical energy requirements of water electrolysis, including the reversible cell voltage and system efficiency. The most common models of electrochemical cells are based on Faraday's law and are used to determine hydrogen production as a function of current, operating time, and Faradaic efficiency. Although these models have a theoretical upper limit, they are typically inadequate for modelling the actual losses caused by temperature, part-load operation, and system aging. Unpredictable power input in PV-driven systems also contributes to the difficulty in characterizing electrochemical behaviours and, consequently, to discrepancies between predicted and experimental hydrogen yields. Such constraints are the reason to combine data-driven correction models to improve predictive accuracy.

2.2.2 Numerical Simulation and System-Level Modelling.

Some numerical techniques used to investigate mass and heat transfer, as well as gas evolution, in electrolyzers include computational fluid dynamics (CFD) and finite element analysis (FEA). These models improve cell design and durability, though they are computationally intensive and not suitable for real-time prediction. Simulation models such as Aspen Plus and the Engineering Equation Solver (EES) have been employed at the system level to study renewable-based hydrogen and power-to-fuel systems, where hydrogen yields are highly sensitive to renewable variability and system efficiency.

2.3 Machine Learning in Renewable Hydrogen Production

The world is moving towards renewable and green energy, and research into sustainable energy solutions is intensifying, especially for producing hydrogen as a clean, green energy source. In the referenced work, the author reports significant progress in exploring the use of Machine Learning (ML) to predict hydrogen production from wind energy in Islamabad, Pakistan [13]. The article is dedicated to the problem of unpredictability of wind energy and its incorporation into the grid and relies on the idea of using Machine Learning (ML) to predict the speed of the wind and estimate the level of hydrogen yield. In their research, the researchers used an NRG 40H anemometer to gather wind data (1 year) in terms of speed, direction and gusts during 10 minutes. Machine Learning three models, including Long Short-Term Memory (LSTM), Support Vector Regression (SVR), and Linear Regression (LR) have been trained to predict wind speed, and hence utilized in predicting hydrogen in a 1.5 MW wind turbine, which corresponded to a Proton Exchange Membrane (PEM) electrolyzer. The LSTM in their research was relatively good. The assumed efficiency of wind turbine was 40 percent and electrolyze was 75 percent. The study had an estimated average of hydrogen production of 6.67kg/day. The findings indicate that there is a correlation between the wind speed and Hydrogen production with high production in the afternoon and higher wind speeds. This technique demonstrates the viability of wind energy in cities. It demonstrates that the application of ML in pre-estimating energy production is useful to reduce the cost of installation and deployment of the turbines.

The author of the mentioned work employed machine learning to optimize the process of producing hydrogen, relying on the progressive photovoltaic (PV) cells technology, thus contributing to a sustainable energy, solution [14]. The study aims at photovoltaic optimization, especially the Passivated Emitter and Rear Cell (PERC) technology, in producing hydrogen. The author gathered data about the PERC system, which reached the output powers of 1, 3, 4, 6, and 7 MW, which suggests the prospects of large-scale hydrogen production. The author has made an exact analysis of the cost of hydrogen levelized (LCOH) with cost savings at a steady 20 percent conversion of 20, 50, and 70 percent of the system component. The prediction of hydrogen production is done using Machine Learning to determine the correlation between solar output and maximum power point tracking (MPPT) output results of three PV technologies, indicating the improved energy production. The paper will point out the possibility of using machine learning to optimize or decrease the cost of making renewable hydrogen.

In the source mentioned, the author has introduced an interpretable deep learning architecture, Optimized Hydrogen Generation-based Regression (OHGR) to predict hydrogen yield in proton exchange membrane fuel cells (PEMFCs) [15]. This model takes into account important input parameters like temperature, pressure and humidity to ascertain the amount of hydrogen produced in different operating conditions. It is a convolutional neural network (CNN) model that incorporates PCA as a dimensionality reduction method. Other than power prediction, researchers in their recent work suggested that direct ML models can be used to estimate hydrogen production. The connections among the renewable power production, the operating conditions of the electrolyzer have been unpuzzled using data-driven models and the environmental variables that ultimately determine hydrogen yield. Interpretable deep learning frameworks and hybrid ML models have achieved higher precision than fully physics-based methods. Of particular importance is that hybrid methods integrating electrochemical processes with ML-based correction terms have been shown to be highly efficient at mitigating real-world inefficiencies [11]. Moreover, Support Vector Regression has demonstrated that

it may have excellent generalization capability for renewable energy datasets in tabular form and data-limited environments, thereby being viable for industrial deployment. Machine learning techniques have been utilized to a great extent in predicting renewable power generation from solar and wind energy sources [13], [17]. Different machine learning algorithms such as Support Vector Regression (SVR), Random Forests (RF), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and XGBoost have shown excellent capabilities in establishing the nonlinear relationships between weather parameters and power generation. To accurately predict renewable power is the first condition to forecast hydrogen in PV–E systems. Using wind data-based hydrogen systems, the results have proved that machine learning (ML) technology to wind speed forecasting can be effectively utilized to raise the production rate of hydrogen and lower the associated investment risks [13], [16]. Furthermore, the solar forecasting model based on GHI, temperature, and irradiance can provide better prediction results than conventional statistical models.
enhancing training efficiency as well as prediction accuracy. To improve the model's performance, stochastic gradient descent (SGD) is employed, and its performance is evaluated using regression metrics, including RMSE (0.220), MAE, and R^2 (0.564), which indicate an improvement over previously published model. This paper concentrates on the increased importance of artificial intelligence in green hydrogen research and the issues of excessive production costs and energy imbalance. By facilitating a data-driven optimization, the OHGR model will optimize the production of sustainable hydrogen and decrease environmental effects.
As stated in the above source, the author concludes that it is possible to substitute grey hydrogen with green hydrogen produced using wind power in China [16]. The study estimated region-specific production costs of hydrogen using real wind variability and eight wind turbines models using high-resolution wind speeds obtained by the use of MERRA-2 database and optimization techniques. This study estimates that wind power-generated hydrogen can replace as much as 28.69 million tons of grey hydrogen by 2040 and 92 million tons by 2060. The sensitivity analysis showed that the key cost drivers are the electrolyzer efficiency and the wind turbine prices. The research lays a more realistic and practical base on the planning of green hydrogen infrastructures, consequently enabling the low-carbon change in China.
The renewable energy has been boosted by the transformation of the world towards renewable energy. Hydrogen production provides a sustainable solution to combating climate change and reducing dependency on fossil fuel resources. In the mentioned reference, the author explain the viability of hydrogen production with the assistance of sunlight in Islamabad, Pakistan, using machine learning (ML) techniques to forecast the output product of a photovoltaic-electrolytic (PV-E) system, which may be utilized as a hydrogen production predictor from renewable energy [17]. The study employs three machine learning models: Prophet, Stochastic Gradient Descent (SGD), and Seasonal. The Auto-Regressive Integrated Moving Average with Exogenous Variables (SARIMAX) model is employed to model Global Horizontal Irradiance (GHI) and resulting hydrogen production. Meteorology Variables

like GHI, air temperature, wind speed, and air pressure. Data were gathered for 13.5 months with every 10-minute observation in Islamabad. Data were divided into a 12-month training dataset and a 45-day test dataset, and relative humidity and deviation in DNI were removed since they had poor correlation. Monocrystalline PV panels and Proton Exchange Membrane (PEM) electrolyzer were simulated, and hydrogen output was calculated as a function of PV electricity output and electrolyzer efficiency. Forecasting models were compared with five other measures of error: RMSE, MSE, MAE, R^2, and MAPE. Prophet performed better than SGD and SARIMAX since it had the lowest MAPE and highest R^2, meaning higher appropriateness and accuracy. The research indicates Islamabad's high solar hydrogen potential, even though the forecast horizon is near winter, proving the feasibility of solar energy in green hydrogen production. In the mentioned reference, the author presented a renewable energy-based system for generating hydrogen and methanol using renewable energy, such as wind energy, and industrial CO_2 to produce methanol [18]. The mentioned system is connected with a wind turbine, electrolyzer, methanol synthesis reactors, and a distillation unit. Electric power from a wind turbine is used in the electrolysis process, and then a portion of the hydrogen produced is used with the CO_2 from industry to make methanol. This system is very efficient as it achieved 42.3% and 40.5% respectively, and it proved to be very useful in reducing CO_2 emissions by about 2999 tons per day. Engineering Equations Solver (EES) and Aspen Plus software are used in this study, which thoroughly examines important variables such wind speed, turbine efficiency, and gas flow rate. The aforementioned study demonstrates how this system functions and the interconnectedness of several elements, including wind speed, energy generation, hydrogen production, and methanol synthesis. These findings will help us make a machine learning model that will help us in forecasting hydrogen.
As the world moves towards renewable and clean energy, integrating machine learning with renewable energy sources shows excellent results, especially in a model where hydrogen production is predicted. In the mentioned reference the author present a detailed study of a model in which photovoltaic (PV) cells coupled with batteries are used to produce hydrogen [19]. Energy coming from PV depends on sunlight, and sometimes sunlight is not even available during the daytime due to clouds and smog. That's why a smart management system is needed to handle production from solar energy. In this proposed system, photovoltaic (PV), battery storage bank, electrolyzer, and fuel cell are used. And for Maximum Power Point Tracking (MPPT), fuzzy logic is used, which performed very well. In the proposed system, solar panels are used to provide power, and the surplus energy is used to produce hydrogen for the fuel cell. In this system a control logic is used, a priority based algorithm that allocates extra energy between battery charging and electrolysis process. As Machine Learning is used in the proposed system, with machine learning prediction of hydrogen production is possible, and can make better decisions to smartly manage energy. Machine Learning has high potential for the prediction of hydrogen.

2.4. Hybrid Physics-Informed and Data-Driven Approaches

Hybrid modeling frameworks combine the first-principles electrochemical models with machine learning predictors to improve the accuracy and trustworthiness of the results. In these models, the hydrogen output calculated theoretically from Faraday's law is adjusted by ML models trained on the historical operational data. The method of using ML models to capture nonlinear losses arising from temperature, degradation, and dynamic loading while still maintaining physical interpretability is the emphasis here. Some recent works utilizing hybrid SVR, RF, and deep learning correctors have achieved substantial reductions in the prediction error; quite often, the annual hydrogen forecast deviations have been brought down to less than 5% in these cases. These experimental results are the direct evidence of a two-stage frame, which is employed to the current research, that PV power is forecasted in Stage 1 and hydrogen production is assessed in Stage 2 by a physics-informed ML corrector which is explained by using Figure 2-4

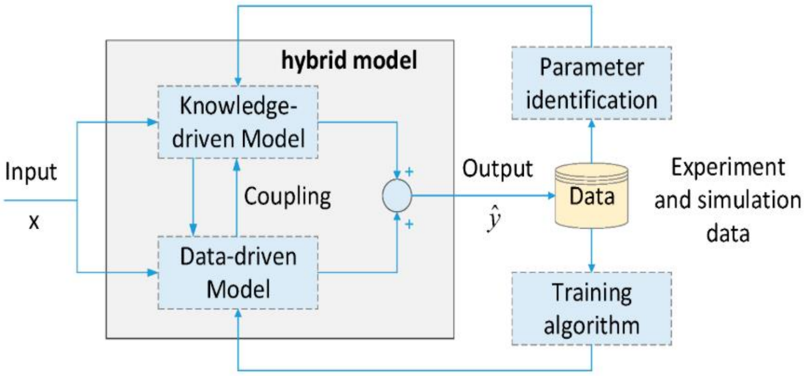

*Figure 2-1 The hybrid model's working mechanism **[20]***

2.5. Sensitivity Analysis and Uncertainty Quantification

The major uncertainties in the renewable hydrogen systems are attributed to the errors in solar irradiance measurements, variation of temperature, and fluctuation of the electrolyzer efficiency. In order to estimate the relative significance of these uncertainties, global sensitivity analysis schemes, including Sobol variance-based schemes, have been embraced over the recent past [21]. The majority of the papers considered mention that the uncertainty in the production of hydrogen is caused primarily by solar irradiance, after which electrolyzer efficiency and thermal effects are observed. The fact that sensitivity analysis is used in the ML-based hydrogen forecasting models further enhances the confidence on the models. It supports sound decisions regarding the theory and approach to the design and running of the system as explained in Figure 2-5 .

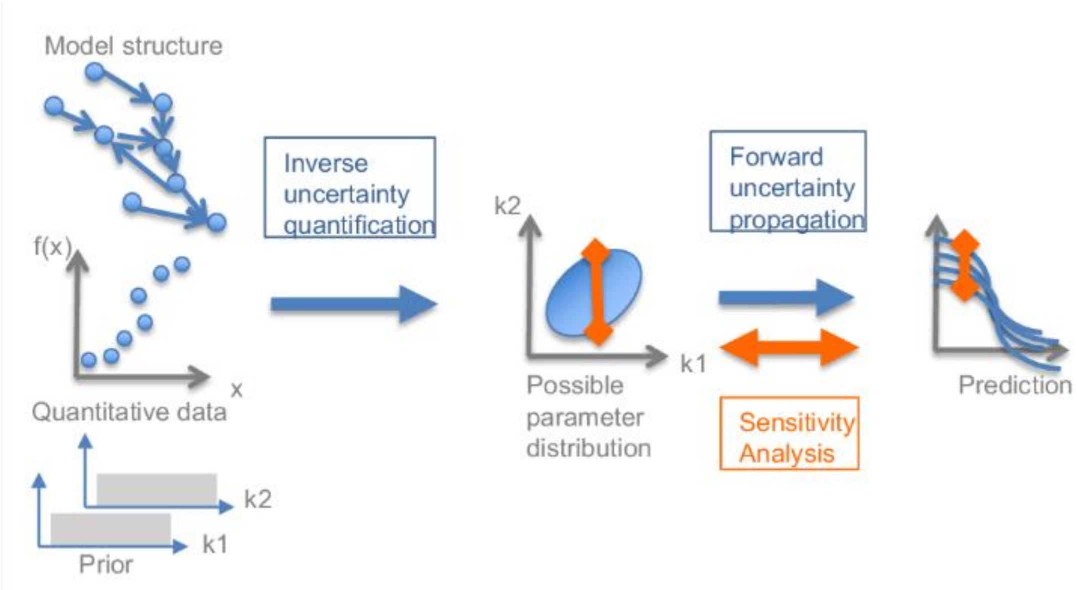

*figure 2-2 Model Sensitivity Analysis **[22]***

2.6. Research Gaps and Motivation for the Present Study

Although more and more studies on the topic of machine learning (ML) -based forecasting of hydrogen production are being conducted, there are still gaps in the existing literature that remain open [23]. A few articles concentrate on the prediction of renewable power only or estimating hydrogen hence ignoring the necessity of incorporating these two phases into one framework. In addition, very little study utilizes high-resolution and utility-scale PV datasets along with physics-informed ML correction models and uncertainty analysis. The present research closes these gaps by presenting a two-step, physically inspired but data-driven model of forecasting solar hydrogen production by utilizing a grid-connected system. The contribution made in this work through models based on SVR, electrochemical theory, and global sensitivity analysis is a good step to having a good scalable solution that is robust enough to address the demands of real world operations.

The literature review provided here reveals the fact that the use of machine learning to create photovoltaic-based hydrogen production systems results in tremendous improvements in the prediction accuracy, system reliability, and optimization. The renewable hydrogen systems with complex nonlinear interactions can be modeled by hybrid physics-informed ML, particularly the Support Vector Regression. The publications reviewed in this context have provided an excellent foundation for the methodology and modelling approach adopted in this study. Although there are numerous studies that estimate renewable power or hydrogen production with the help of machine learning, there is very little effort put into creating a physics-informed and two-step model that connects the ideas of high-resolution photovoltaic power prediction and electrochemical hydrogen modelling with uncertainty analysis and operational feasibility in a real-world context. This is the primary reason behind the approach applied in the present work.

Table 2-1Comparative Analyses of Research Reported in Literature

[Ref.], (Year)	Problem	Machine learning / Modeling Approach	Results and limitations
[13], (2022)	Hydrogen production from wind energy	LSTM, SVR, Linear regression	LSTM algorithm performed better than the other two
[14], (2024)	Prediction of hydrogen production using advanced photovoltaic technologies (PERC)	Methods used are regression or advanced ML models like Random Forest or Neural Networks, based on context.	A Levelized Cost of Hydrogen (LCOH) sensitivity analysis showed cost reductions of 20%, 50%, and 70% at 20% conversion rates.
[15], (2024)	Prediction of hydrogen generation in PEM fuel cells using temperature, pressure, and humidity	Deep Learning (CNN), PCA, Stochastic Gradient Descent (SGD), SHAP	SHAP performed best among them all.
[16], (2023)	Substitution potential of wind-powered hydrogen production (WPHP) for fossil-based hydrogen in China	Mixed-Integer Linear Programming (MILP) optimization	WPHP has strong potential to replace grey hydrogen in various regions; electrolyzer efficiency and wind turbine cost were the most influential factors
[17], (2021)	Forecast of the production of hydrogen, using solar energy in Islamabad through water electrolysis	Prophet, SARIMAX, Stochastic Gradient Descent (SGD)	Prophet performed best among the three. With Great results.
[18], (2020)	A system was proposed to produce hydrogen and methanol using wind energy and industrial CO_2	The analysis was performed without the aid of machine learning using Engineering Equation Solver and Aspen Plus for simulation.	Production was as follows: 1.13 g/s hydrogen and 52.25 g/s methanol; reduced 2999 tons/year of CO_2; efficiency ~40%

[19], (2020)	A standalone photovoltaic/battery/hydrogen system was proposed to manage the energy needed to supply a house in Algeria.	Fuzzy Logic Controller (FLC), Perturb & Observe (P&O), Incremental Conductance (Inc) were used.	FLC was best: fastest response, most stable power, and best at using sunlight efficiently.
[9], (2017)	Solar power generation forecasting	Support Vector Regression (SVR), Artificial Neural Networks (ANN)	ML models significantly improved solar power prediction accuracy; hydrogen production not considered
[10], (2017)	Solar irradiance forecasting for renewable energy systems	Review of ML techniques (ANN, SVR, hybrid models)	Demonstrated superiority of ML over statistical models; study limited to irradiance prediction only
[11], (2013)	Performance modeling of PEM water electrolysis	Electrochemical and thermodynamic modeling	Accurate modeling of electrochemical losses; lacks data-driven prediction under variable renewable input
[12], (2018)	Dynamic behavior of PEM electrolyzers under fluctuating power	Dynamic electrochemical modeling	Highlighted strong sensitivity to power variability; real-time hydrogen forecasting not addressed
[1], (2023)	Global energy transition and role of green hydrogen	Policy analysis and scenario-based modeling	Identified hydrogen as a key decarbonization vector; no system-level or ML-based modeling
[18], (2019)	Integrated hydrogen and methanol production using wind energy	Aspen Plus and Engineering Equation Solver (EES)	Achieved ~40% system efficiency and significant CO_2 reduction, no machine learning integration

Chapter 3
Background

3.1. Solar production:

Photovoltaic cells are the simplest components of photovoltaic systems. Semiconductors are the main categories of photovoltaic cells, which serve to conduct photoelectric effect. The photovoltaic cell also known as solar cell is a process which involves photovoltaic cell which is an electronic device that directly transforms light energy (photons) into electrical energy (electrons) by the photovoltaic effect. Silicon is an element that is widely used in making up mainly photovoltaic cells, since it is abundant, stable and suitable.

The processes of the photovoltaic cell constitute various stages of semiconductor production.

The steps include:

- First step would be to select regarding the materials which are used to make photovoltaic cells such as crystalline silicon, cadmium telluride and gallium arsenide.
- Second step is the doping step, to form an electric field within the semiconductor. The material is separated into two areas like n-type and p-type where p type is the carrier of positive charges doped with phosphorus and n type is the carrier of negative charges doped with boron.
- The third step is to make layers of these two areas the layers primarily consist of the front contact that is a fine metallic grid that gathers electrons. Following that anti-reflective coating assists in preventing the process of reflection to allow the absorption of the greater number of photons.

Finally, one n-type region layer and one p-type region layer are followed by finally back contact which is composed of silver material to offer final layer.

At this point the principle of photovoltaic effect is incorporated to combine the layers which produce the voltage at the same time produce the current upon exposure to sunlight since the sunlight consists of small packets of energy commonly referred to as photons when combined with the semiconductor material they are absorbed and further lead to the excitation of electrons in the n type region. These electrons that tend to reside in the valance band then tend to flow through the conduction band as a result of the 4excitation. This excite forms electron-hole pairs [23]. At this point when the p-n junction is composed of electric field, then, the electric field drives the electrons into the n-type region and the electrons called protons into the p-type region. This isolation confines the electron and protons to come back together and a potential difference creates (also referred to as voltage in the electrical world). The last working involves an external circuit of the photovoltaic cells and because of this the electrons begin to flow through the load, between the n-type region and the p-type region producing a direct current (DC). The mathematical expression that represents the electricity or energy that is produced through the work can be represented as:

$$E = \frac{hc}{\lambda} = hv \quad (1)$$

Where E is the photon energy, h is the Planck constant, c the speed of light and λ is the wavelength as illustrated in the above equation. The weaknesses of these photovoltaic cells are that they are expensive to build and the amount of electricity that is produced is at the mercy of sunlight.

Solar photovoltaic systems move with the movement of the sun. The sun is not always constant, so the output of a panel will be jagged at the same time of the day and the same time of the year. The quantity of solar radiation reaching a panel varies with the time of the day in the rising and setting of the sun. The sun is also low early in the day and late in the day and in each case casts its light at an angle, providing less intense radiation, which cuts off the electrical output. With the approach of noon, the sun is gaining height, the angle of incidence is decreasing and the panel is able to receive more direct sunlight. In the solar noon, the irradiance is at its highest/peak and the PV system normally reaches its maximum power [24]. Subsequently, the intensity diminishes and the energy production declines as the sun slides downwards creating the same bell-shaped daily curve characterising the daily motion of the sun.

Tilted fields Seasonally the playing field is tilted by the tilt of the Earth and the orbit. The increased arc of the sky in summer days and the increased length of time (summer) that the sun is caught increases the amount of sunlight available to the complete span, and a higher solar altitude increases the direct radiation and thus efficiency and overall output. Winter countermeasures this: the number of days and the path of the sun become shorter and the amount of energy that can be produced decreases, and the energy production decreases. In between are spring and autumn, which have moderate radiation levels and output, and are periods of transition between the abundance of summer and the scarcity of winter. At the same time, cycles emphasize the fact that the production of solar energy highly depends on time and is determined by the foreseeable motions of the Sun [25].

Analysis of PV system behavior on different time scales depicts that good energy capture by the photovoltaic involves adaptive control systems that react to temporal change in the irradiance. Time, then, proves to be the most significant in the process of solar energy generation both in the short-term (daily) to the long-term (seasonal) nature of its output, and demands an efficient approach of managing it to ensure its optimal production in natural varying conditions.

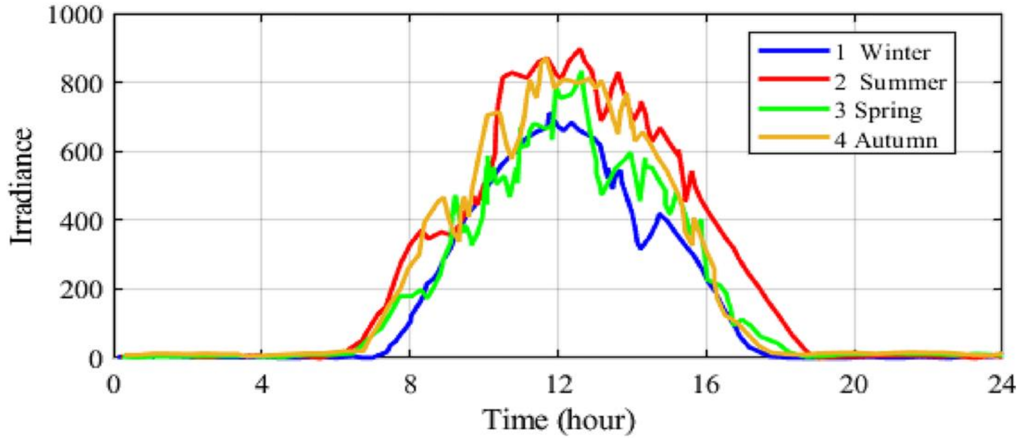

Figure 3-1 Irradiance levels in Hong Kong in different seasons [26]

Energy coming from solar will power the electrolysis, as shown in the following equation (2)[27].

$$P_{PV} = A \times G \times \eta_{PV} \times PR \quad (2)$$

Where, P_{PV} is the power output of the module (W), A is the surface area (m²), G is solar irradiance (W/m²), η_{PV} is PV module efficiency, and performance ratio (PR) is the module's Efficiency.

Now, the mathematical calculation for solar irradiance is mentioned by equation (3):

$$G = I_0 \cos\theta_z \, e^{-\tau} \quad (3)$$

(4)

Where, I_0 is the solar constant (1361W/m2), θ_z is solar zenith angle, and τ is the atmospheric optical depth.

3.2 Electrolysis:

Hydrogen is a clean, multi-purpose and highly efficient fuel; it can fuel the fuel cells, propel cars, aid the industrial processes, and even produce power without carbon dioxide. It is relevant because of the reasons, including the fact that it preserves energy derived through renewable energy sources like solar and wind. It can also be used to decarbonize steel, cement, and chemicals sectors of the heavy industry to allow fuel-cell cars to transport and allow mobility, stabilize the grid, and allow a storage of energy in a seasonal manner [28].

Hydrogen is the lightest but most common element in the universe though it is not found in large quantities freely in nature. It typically occurs in complex with other substances, e.g. oxygen in water (H_2O) or carbon in hydrocarbons (CH 4, coal, petroleum). Hydrogen is produced using these compounds with the assistance of chemical or electrochemical reactions.

The preparation of hydrogen can be done in a number of ways as discussed below:

1. Hydrogen made using fossil fuels (Grey and Blue Hydrogen)

2. Hydrogen made out of biomass (Bio-hydrogen).

3. Hydrogen through hydrogen electrolysis over water (Green Hydrogen).

The manufacture of green hydrogen is done under a process called electrolysis of water whereby the energy used in the process is a renewable source such as solar, wind, or hydropower to split water into hydrogen and oxygen gases. This begins by the production of clean electricity through renewable sources and then it is fed into an electrolyzer the main unit that separates the molecules of water. The pure or deionized water is added to the electrolyzer and it has two electrodes that are the anode (and) and the cathode (that is, negative, so that the gases do not mix). The use of direct current gives the energy needed to break the bonding forces existing between hydrogen and oxygen atoms in water. Due to this reason, the water molecule dissociates: the hydrogen ions move towards the cathode, where they acquire electrons and join to form hydrogen gas; the hydroxide ions or water molecules dissociate at the anode to release electrons and form oxygen gas[29].

The overall chemical reaction is as follows:

$$2H_2O(l) \rightarrow 2H_2(g) + O_2(g) \tag{4}$$

This implies that two water molecules produce two molecules of hydrogen and one molecule of oxygen. The hydrogen gas that is generated at the cathode is collected, purified and then stored under pressure in appropriately designed tanks or pipelines to be used later. The one by-product of the process is oxygen that goes to the atmosphere or is taken to be used in industry and medicine.

The green Hydrogen efficiency and nature of production varies according to the type of electrolyzer used. Moderate temperatures are used in alkaline electrolyzers, which use a liquid electrolyte, usually KOH. Conversely, electrolyzers based on PEM utilize a solid polymer membrane, which permits the flow of hydrogen ions, and thus yields very pure hydrogen with fast response times. Solid oxide electrolyzers are at high temperatures and can operate either on electrical or thermal energy which allows them to be used in very large-scale and high-efficiency systems. The electricity that provides power to such electrochemical electrolyzers is usually produced using renewable energy technologies, as in the case of solar PV panels or wind turbines. Since during the stanzas of high renewable energy production, the electrolyzer will receive excess electricity, it will unquestionably be feasible to charge the hydrogen manufacturer so that, in the long-term, a continuous production of hydrogen is sustained without any increased greenhouse gas emissions into the atmosphere, making green hydrogen manufacturing one of the most viable and environmental friendly methods of generating fuel. In addition, it also helps in achieving global carbon neutrality by substituting grey or blue hydrogen, which is generated using fossil fuels, and helps in deep decarbonization in most sectors intensive in energy.

Electrolysis of water to produce green hydrogen, thus, represents one of the major technological directions towards a cleaner, more resilient and sustainable energy future.

The electrolysis process is represented mathematically in equation (5) by using Faraday's law .

$$H_2\left(\frac{Nm^3}{h}\right) = {I.t}/{n.F} \tag{5}$$

Where I is current (A), t is time (seconds), n is Electrons per H_2 molecule, and F is the Faraday constant (96485 C)

Chapter 4
Methodology

This study constructs a decision making, two-stage machine learning model to make predictive modelling and operational forecasting of a grid-tied solar hydrogen production system. The process will transform high-rate weather and photovoltaic (PV) system data into useful forecasts of power output and subsequent generation of green hydrogen.

The system switches to the 15-minute-interval weather and photovoltaic (PV) system-data in a gradual approach so that it can be utilized to make predictions of both the power output and the ensuing hydrogen yield, which can be easily used.

This study has an empirical foundation of the time-series data that was gathered at the 40.5MW photovoltaic system in Chikkoppa village, Karnataka, India (15.652016o N, 75.992484o E) [34]. The data, acquired in an authenticated Kaggle repository, is composed of concomitant measurements on the 15-minute levels throughout the multi-year time span that the plant was in operation. The data set that is taken to model hydrogen production in this study comprises of the meteorological and electrical parameters. The meteorological parameters are ambient temperature (AT, C), module temperature (MT, C), and global horizontal irradiance (GHI, W/m 2) and the electrical parameter is the AC power output (AC POWER, kW). The dependent variable is hydrogen production (H 2), in kilos and per 15 minutes. A number of data conditioning procedures were being used prior to developing the model in order to guarantee quality and consistency of the data. To do this, first column standardization was done by automated feature name detection and harmonization of feature names across data variants. Second, the zero-production periods were screened with the observations with hydrogen production less than 0.1 kg being eliminated to enable the model to concentrate on the plausible operational conditions. Third, the analysis of missing data was done through complete case analysis where all the empty records were excluded. Lastly, the temporal segmentation was used to maintain the chronological organization of the data to time-sensitive modelling. The filtered dataset was subsequently subdivided chronologically into two non-overlapping segments or subsets, a training set (80 percent of the data or approximately 20182020) and a test set (20 percent of the data or approximately 2022).

The model of hydrogen production that has been designed in this study is based on the features of a commercial system of MW-scale Proton Exchange Membrane (PEM) electrolyzer. The electrolyzer has a fully automated, containerized design, with tri-mode capability, to allow command-following, load-following, and tank-filling operational modes, so that it is highly suited to the integration of variable renewable energy. The Hydrogen production can be dynamically regulated between wide operating range of 10% to 100% of rated capacity, rapid system response is listed with start-up times taken less than 5 minutes and ramp rates more than 15 percent of full-scale power per second. The electrolyzer has an average specific energy consumption of about 50.4 kWh/kg of hydrogen, which gives a solid physical correlation between the AC power input and the hydrogen output that can be used to model it with data. Purities of hydrogen are available up to 99.9995% (including purification) and the nominal pressure of delivery is 30 barg, which guarantees the production of industrial grade. To make sure that the machine learning-based forecasts of hydrogen production are physically consistent and reflective of the actual conditions of operation of the machine, operational constraints, such as ambient temperature limits (−20 o C to 40 o C), electrical power quality requirements, and the rates of deionized water consumption, were taken into account. Daily and monthly production of AC Power and hydrogen of the site are displayed by using graph in figure 4-1, 4-2 4-3 and 4-4.

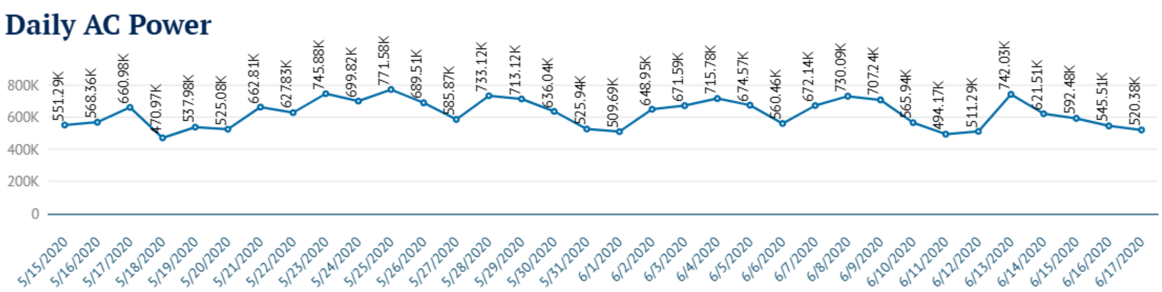

Figure 4 -1 Daily Ac power production from solar

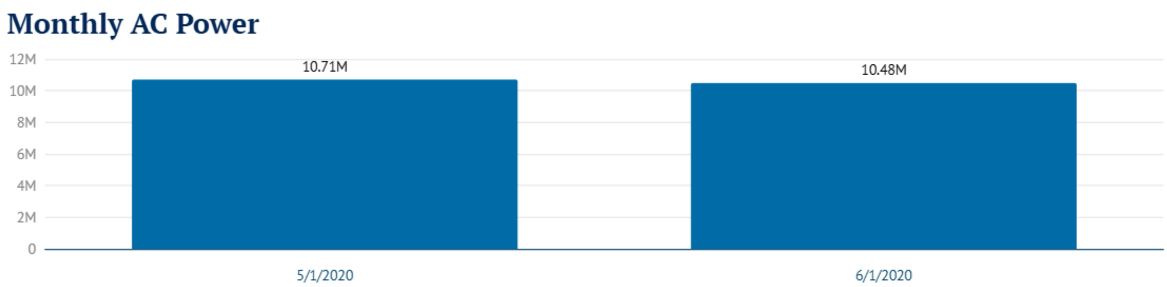

Figure 4 -2 Monthly Power Production

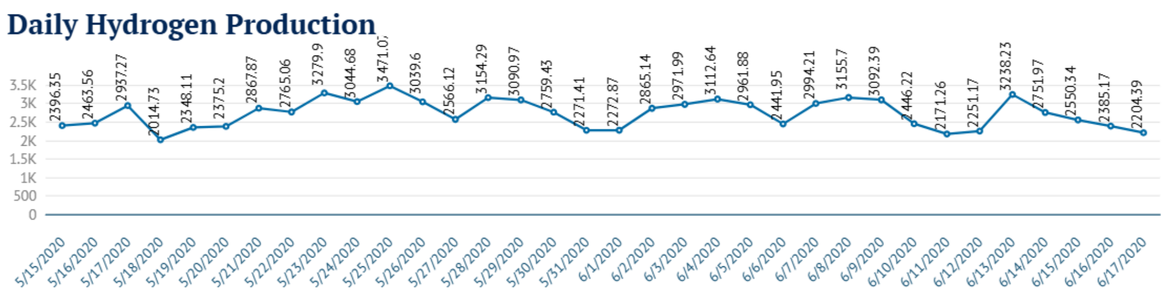

Figure 4 -3 Daily hydrogen production

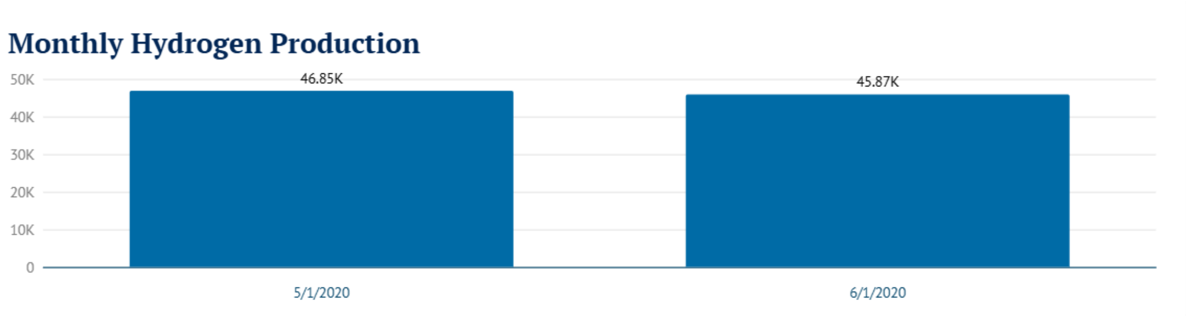

Figure 4 -4Monthly Hydrogen Production

4.1 Machine Learning and Deep Learning Models

Machine Learning (ML) is a core branch of artificial intelligence that enables computer systems to learn from historical data and improve their performance without explicit programming. The ML algorithms detect the presence of hidden patterns, trends and relationships in data, by optimizing mathematical objective functions. The learning process is generally accompanied with the data collection, preprocessing, feature extraction, model training, validation and testing. Machine learning methods are generally divided into supervised, unsupervised, and reinforcement learning, depending on the type of the learning task. Training models on regression and classification problems in supervised learning involves the use of labeled datasets, and deep learning models work with multi-layer neural networks in order to learn and capture complex nonlinear interactions and temporal correlations.

This paper discusses both traditional machine learning algorithms and deep learning neural networks with a view to evaluate their prediction power. The models of choice are Decision Tree, Random Forest, Gated Recurrent Unit (GRU) and Support Vector Regression (SVR), the latter being the proposed model because it has high generalization and theoretical strengths.

4.2 Decision Tree

A Decision Tree is a supervised learning algorithm which represents the process of making a decision in the form of a hierarchical tree, with the internal decision nodes, branches and terminal leaf nodes as the nodes of the tree. A decision rule represented by each internal node, which is determined by a particular input feature with each branch representing the result of the decision, and the predicted values of the output are also stored in the leaf nodes. Decision Trees recursively split the input feature space into smaller and more homogeneous subsets with the aim of reducing the error of prediction.

In classification tasks there are commonly used metrics like entropy and information gain which are used to identify optimal splits. On the contrary, with regression problems, the splitting criterion depends on the reduction of variance. The difference of a dataset is referred to as variance.

$$\text{Var}(S) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{1}$$

where y_i is the target value, and $\bar{y}$ is the mean of target values [35]. Regression trees improve the quality of predictions they make in each region by choosing splits that reduce the amount of variance. The Decision Trees can model nonlinear relationships and they include interactions with minimum data preprocessing. They are however, likely to overfitting especially with an increase in the tree depth, and this inspires the application of ensemble-based methods.

4.3 Random Forest

Random Forest is an ensemble learning methodology which aims to address the weaknesses of the single decision trees by aggregating the outputs of the many trees. It is based on the concept of bootstrap aggregating (bagging) in which the trees are trained on random samples of the training data with replacement. Moreover, a random sub-set of features is chosen whenever the node splits, and this creates variability across trees and lessens correlation.

In regression problems, the overall prediction of a Random Forest model is the mean of all the individual tree predictions which can be written as

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t (x) \tag{1}$$

where T denotes the total number of trees and $h_t$is the prediction of the tree [36]. Such an aggregation process leads to significant model variance reduction as well as robustness as well as generalization performance. Random Forest models are especially useful in high dimensional data and in modeling nonlinear data in a stable environment even under noisy analysis.

4.4 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a recurrent neural network that is suitable to model sequential and time-dependent data. As opposed to the classical recurrent neural networks, GRU addresses the vanishing gradient problem through the implementation of gating mechanisms that control the flow of information in the network.

The update gate is defined as

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{1}$$

which controls the amount of past information retained [37]. The reset gate is expressed as

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{2}$$

and determines how much historical information is forgotten. The candidate hidden state is computed as

$$\tilde{h}_t = \tanh\left(W_h x_t + U_h(r_t \odot h_{t-1})\right) \quad (3)$$

and the final hidden state is updated using

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

GRU networks have been shown to have gating operations that result in a computationally efficient gating of temporal dependencies across long durations [38]. GRUs are also useful in time-series forecasting and sequential prediction because they have very simple architecture compared to Long Short-Term Memory (LSTM) networks.

4.5 Support Vector Regression (SVR) – Proposed Model

In this research, Support Vector Regression (SVR) has been chosen as the proposed model because of its good theoretical basis and good generalization ability [39]. The principle that underlies the use of SVR is the structural risk minimization, which attempts to trade off the complexity of the model and its prediction accuracy. The aim of SVR is to approximate a regression function.

$$f(x) = w^T x + b \quad (1)$$

that is as flat as as possible and does not prohibit deviations within an e-insensitive margin. To obtain a model flatness, the weight vector is minimized and thus helps to prevent overfitting.
The optimization problem is represented as.

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \quad (2)$$

where C is the regularization parameter that regulates a trade-off between the smoothness and the error of prediction, and is a slack variable that is the deviation beyond the e-insensitive margin. SVR uses kernel functions to obtain nonlinear relationships. The Radial Basis Function (RBF) kernel is employed in this paper:

$$K(x_i, x) = \exp\left(-\gamma \parallel x_i - x \parallel^2\right) \quad ()$$

where γdetermines the influence of individual training samples.

To make sure that the proposed SVR model evaluation is reliable and comprehensive, several quantitative measures are employed to assess the performance of the proposed model. The first measure of the error in prediction is the Mean Squared Error (MSE) which quantifies the mean squared error between the actual and predicted value and penalizes a large error more severely. MSE is defined as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (4)$$

where $y_i$represents the actual target value, $\hat{y}_i$ is the predicted value, and n is the total number of samples. The smaller the MSE value, the higher the predictive accuracy and higher the model performance.

The error in the target variable can be expressed in the same units as the error, so the Root Mean Squared Error (RMSE) is used, which is expressed as,

$$\text{RMSE} = \sqrt{\text{MSE}} \qquad (5)$$

RMSE is a more interpretable measure of prediction error and that it emphasizes the dispersion or variation of residuals thus it is especially helpful in comparing the performance of different models on different datasets. Also, the coefficient of determination R 2 is utilized in order to determine the effectiveness of the SVR model to explain the variability that exists in the observed data. It is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (6)$$

where $\bar{y}$denotes the average actual target values. The nearer the R-square is to one the better the model is in capturing the underlying pattern of data in the data and it has a strong explanatory power.

To classify the data, classification-based evaluation is performed based on the precision, recall and F1-score that are obtained based on the confusion matrix. A confusion matrix is a table format that compares the actual labels of the classes with the predicted labels and is made up of four elements which are True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). This matrix gives a finer understanding of the kind of prediction errors that the model has given.

Precision is a measure of the percentage of the positives which the model correctly predicts out of all the positives which the model makes a prediction. Recall measures the capability of the model to represent all real positive cases, which is the sensitivity of the model. F1-score, which is defined as a harmonic mean of precision and recall, is a balanced measure of performance especially in the event of class imbalance. The metrics based on confusion matrix provide a more in-depth insight into the model behavior, as each of them shows the equilibrium between the correct prediction and misclassifications. All these evaluation measures will guarantee a holistic and strict evaluation of the proposed SVR model.

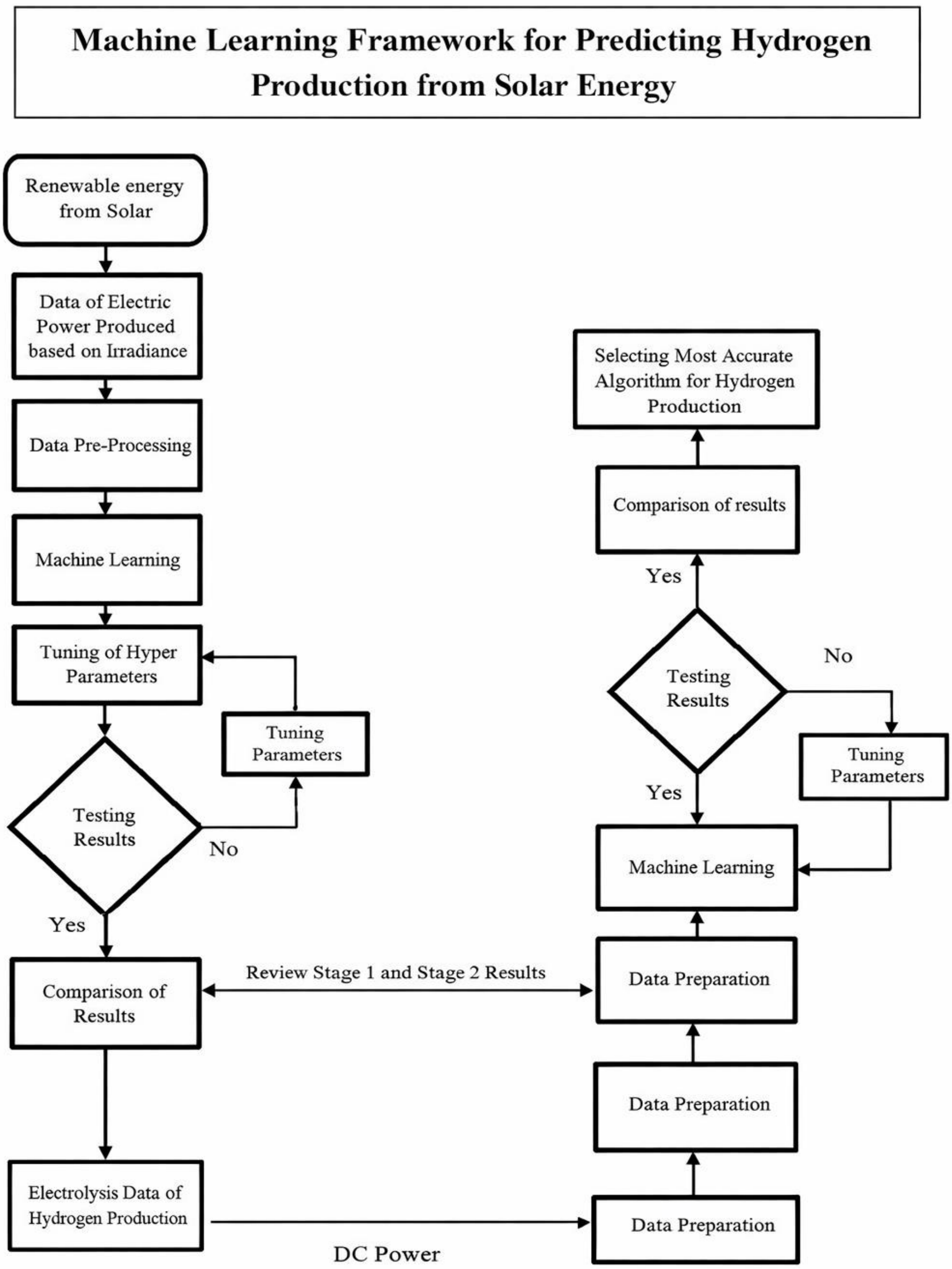

Figure 4 -5 Flow Chart

Chapter 5
Results and Discussion

This research project has presented an overall analysis of the possibility of utilizing excess solar and wind power to produce hydrogen. The quantitative analysis shows the effectiveness of the proposed system under changing generation conditions, where energy production and hydrogen yield are at maximum. These findings not only demonstrate how well hydrogen can serve as a storage medium for intermittent renewable energy but also highlight its potential applications in transportation and industrial sectors. The findings also provide an idea of how the systems can be optimized, namely, in which states the maximum energy use can be achieved. Wastage can be minimized.

5.1 Stage One ML on Hydrogen prediction

5.1.1 Random Forest Regressor

Random Forest Regressor has been used to forecast AC power output depending on the chosen input features. Random Forest is an ensemble method of learning which is used to build many decision trees to enhance accuracy of predictions and minimize overfitting based on averaging of the output of each tree. This ensures that it is specifically effective in modeling non-linear relationships as well as modeling complex interactions between the input variables.

Mean Absolute Error (MAE), root mean square error (RMSE), and the coefficient of determined (R 2) were used to assess the performance of the random forest model. The model had an MAE of 600.26 W which means on an average, the predicted AC power is off the actual values by an average of 600 W. RMSE was of 1963.58 W which shows that it has some higher-order prediction errors, but they are acceptable taking into account the broad scope of the operating range of AC power output. Also, the model gave a high R^2 of 0.9465 which proves that the Random Forest model has described nearly 94.65 percent of variance in the actual AC power output. Figure 5-1 shows the correlation between the measured and predicted values of the AC power of the Random Forest Regressor.

The scatter points all align closely along the 45-degree reference line and it shows a perfect fit of prediction. The high concentration along the ideal line shows that there is a high degree of concurrence between the predicted and actual values at low power range, medium power range and high power range. Slightly dispersed higher power levels imply that there is a slight under or over-prediction at peak operating levels, as is normal in practical power generation data. In general, the Random Forest Regressor has a high predictive power, high robustness, and high generalization. Its large R 2 value and not so large MAE prove its appropriateness to the prediction of AC power and its right to be one of the main models used in this project. The findings show that Random Forest is useful in capturing the underlying nonlinear patterns in the data.

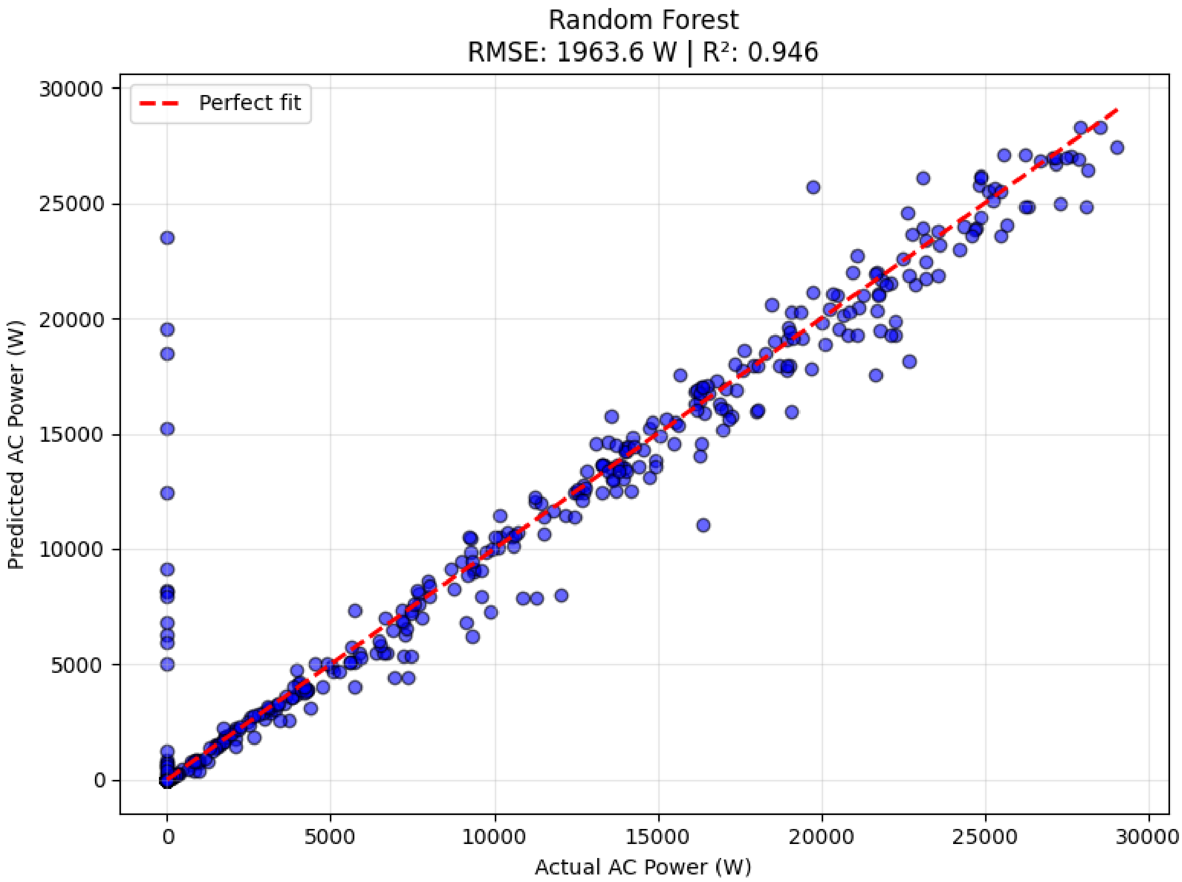

Figure 5- 1 Random Forest model prediction

5.2 Decision Tree Regression:

The Decision Tree Regressor was applied with the purpose of serving as a control machine learning model to forecast the output of AC power on the basis of the clean dataset. Decision Trees are non-parametric intuitive models that divide the input space using feature thresholds, but they are also known to be susceptible to noise and also overfitting with default parameters.

The evaluation of the Decision Tree model was in terms of Mean Squared Error (MSE), Root Mean Square Error (RMSE), and coefficient of determination (R2). The model had an MSE of 5,914,992.15 W 2, which is an RMSE of 2432.08 W, and it means that the variation in the model has a greater average error of prediction than that of the ensemble-based models. The R 2 value obtained, 0.9179, indicates that the model is able to explain a large portion of the variance in the actual power output of AC and on this basis it has a reasonably good predictive potential but has significant weaknesses. The Scatter plot of actual and predicted AC power values of the Decision Tree Regressor are shown in Figure 5-2. Although, there is a general inclination that can be seen and most of the points are near the reference line of 45 degrees, we can see a greater level of a dispersion especially in high power levels. A number of outliers and sudden changes of direction away form the perfect fit line suggest that the model cannot easily generalise across the entire operating range. This is typical of single decision tree models, which are rather sharp and piecewise than smooth approximations. Generally speaking, the Decision Tree Regressor has acceptable base performance to predict AC power but its greater RMSE and smaller R2 than the Random Forest model indicate its low level of generalization. These findings indicate that though the Decision Trees is able to represent simple nonlinear association, ensemble procedures like the Random Forest are more robust and accurate at a lower level of variance and higher predictive stability.

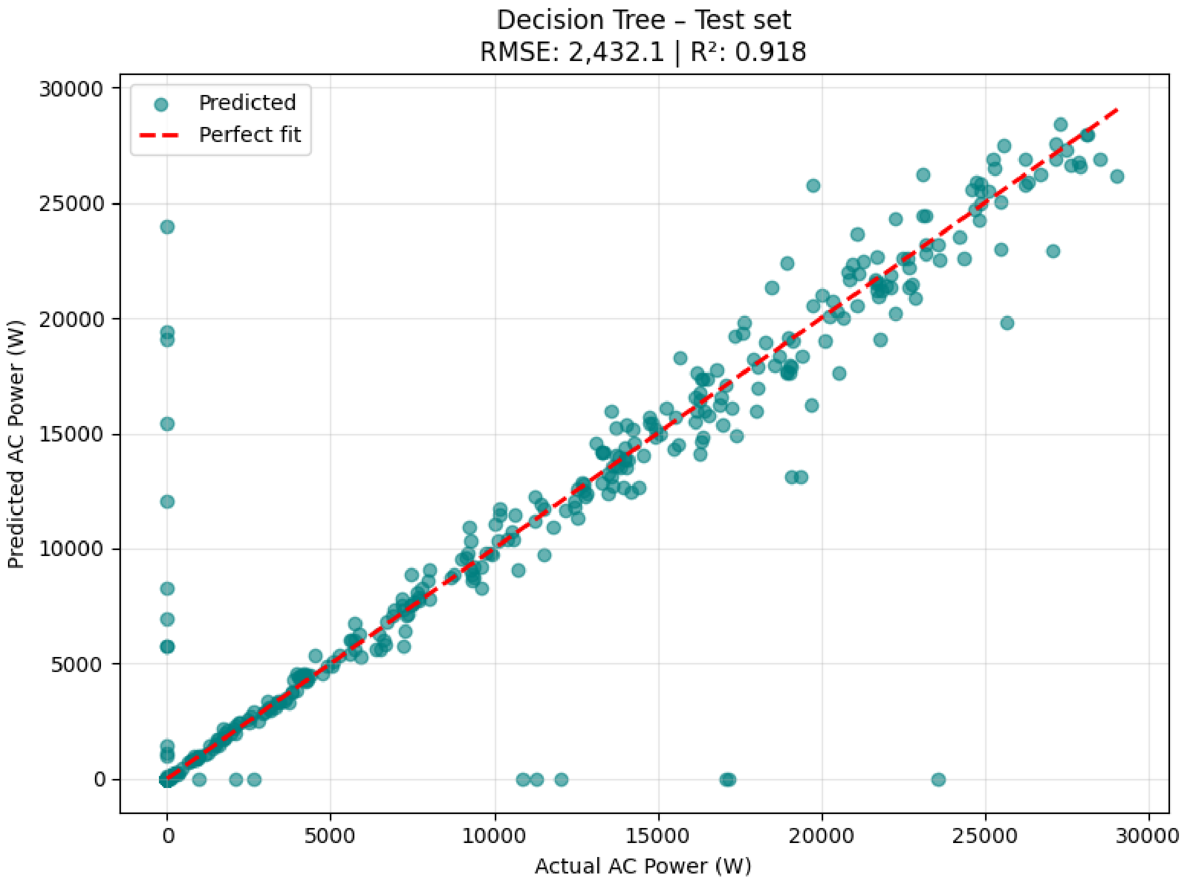

Figure 5-2 Decision tree regression prediction vs. actual production

5.3 GRU Gated Recurrent Unit

GRU model has been used to represent the time-varying values of the AC power generation data. GRU is an RNN based architecture that addresses the vanishing gradient issue through the use of update and reset gates thus enabling the model to learn both short and long term patterns processing. This renders GRU especially applicable when it comes to power prediction applications in which sequential and time-varying variations affect system output. Mean Squared Error (MSE), Root Mean Square Error (RMSE), as well as the coefficient of determination (R 2) were used to measure the performance of the GRU model. The model gave an MSE of 6,157,151.15 W 2, which translates into an RMSE of 2481.36 W 2, and thus, an average degree of prediction error in the test set. The R 2 value of 0.9010 indicates that the GRU model captures about 90.10 per cent of the variance in the actual AC power output.

The comparison between the actual and predicted value of the AC power obtained at the use of the GRU model is depicted in figure 5-3. The ideal 45 degree line is followed by the predicted values and the prediction shows that the model is able to learn the general tendency of power generation. Nevertheless, the dispersion is observed to be greater especially at high power levels, indicating low accuracy under peak operation conditions. This behaviour shows that although the GRU can be useful at capturing the temporal trends, it might need more architectural fine-tuning or input features to make it more precise.

On the whole, the GRU model can be said to have a reasonable predictive performance, with the ability to model time-dependent behaviour in AC power generation. Nevertheless, the GRU has a lower RMSE and a lower R2 than the tree-based ensemble methods like the Random Forest, which implies a lower accuracy. These findings indicate that in spite of the fact that GRU can also learn sequential patterns, the methods of ensemble regression are still more suitable in the context of the current dataset and feature architecture.

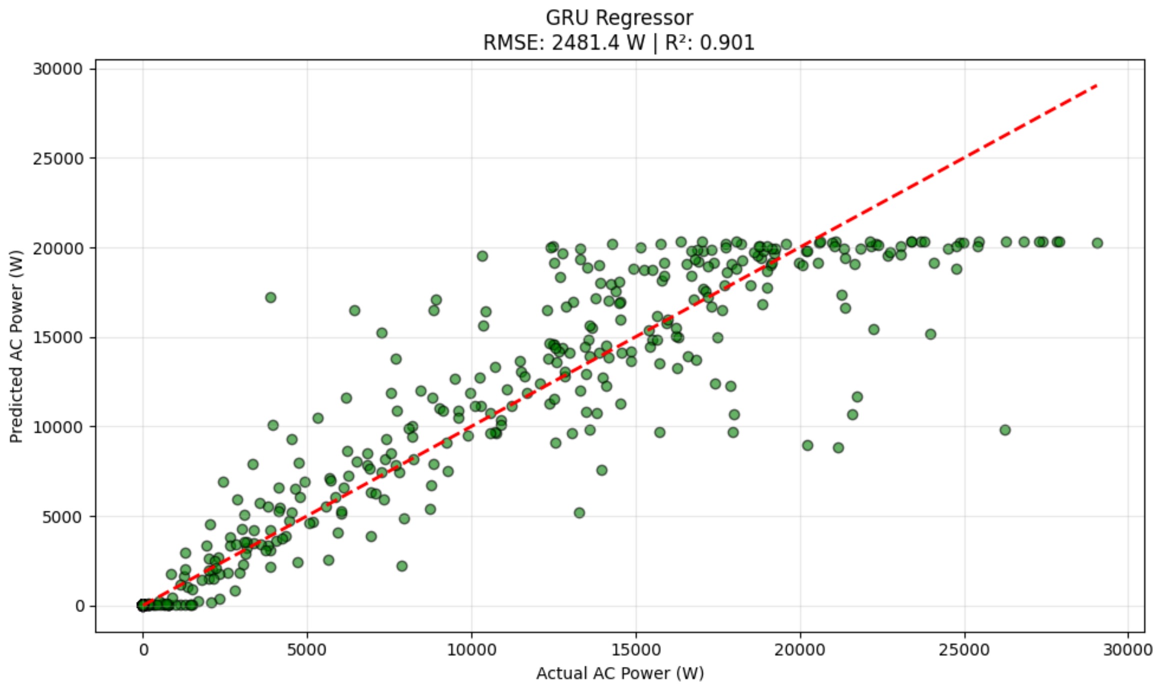

Figure 5-3 Prediction vs. actual production

5.4 Support Vector Regression:

Support Vector Regression (SVR) was utilized to forecast the AC power output as it has a strong capability of modeling nonlinear relationships through kernel, based learning. Essentially, SVR tries to determine the best regression function by minimizing the prediction error within a certain margin (, insensitive loss), at the same time preserving the model complexity via regularization. The trade, off between precision and generalization is what makes SVR a tool of choice in solving complicated power prediction problems.

The efficacy of the SVR modeling was measured in terms of the Mean Squared Error (MSE), the Root Mean Square Error (RMSE), and the coefficient of determination (R). The SVR model optimized to local condition settings resulted in an MSE of 3, 431, 649.60 W, hence a relatively low RMSE of 1852.47 W, pointing to a considerably better prediction accuracy than the Decision Tree and GRU models. Besides, the model reached a very high R value of 0.9523, which means that the SVR model explains about 95.23% of the variance of the real AC power output.

Figure 5, 4 shows a scatter plot of actual versus predicted AC power values for the SVR model. The predictions are very close to the ideal 45, degree reference line, and thus, we can say that there is a strong correlation between the predicted and actual outputs in the whole operating range. Very few outliers can be accounted for at high power, and also, the dispersion is almost negligible.

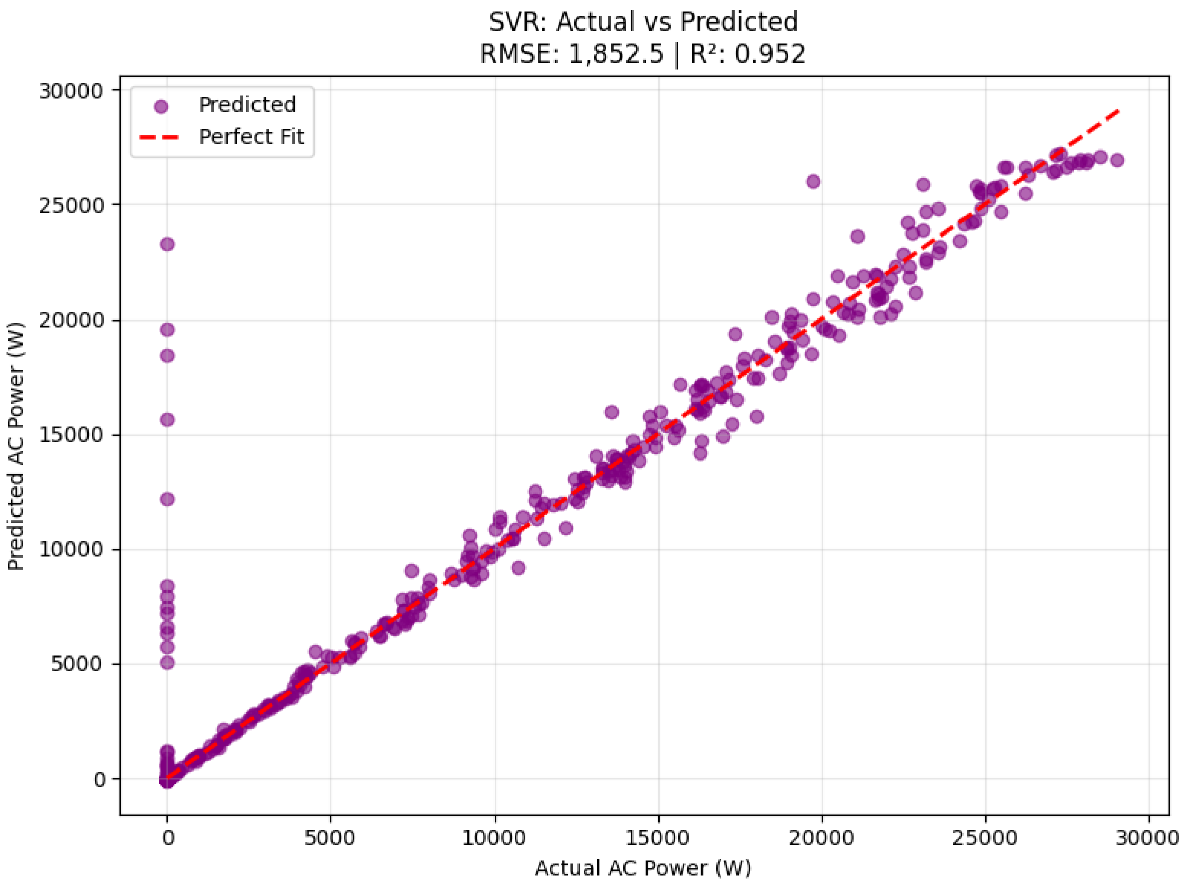

Figure 5-4 Prediction vs. actual production

The performance of different machine learning models for predicting the target variable was first discussed in Table 5, 1, and the corresponding results are summarized in Table 1. The models tested were SVR, Random Forest, Decision Tree, and GRU. Among these, SVR was found to be the best performer, having the lowest RMSE of 1852.47 W and the highest R of 0.9523, which means that its predictions were the closest to the actual values and it explained the most variance in the data. Random Forest performed a little bit worse, followed by Decision Tree and GRU, which had the highest error and lowest R. These findings confirm that SVR is the most precise and dependable model for this dataset

Table 5-1 Comparison table of ML results

Model	RMSE (W)	R^2
SVR	**1852.47**	**0.9523**
Random Forest	1963.58	0.9465
Decision Tree	2432.08	0.9179
GRU	2481.36	0.9010

5.2 Stage TWO ML on Hydrogen prediction

5.2.1 Random Forest Regression Performance Evaluation for Hydrogen Production

Random Forest Regression (RF) was used and used to predict hydrogen production of the solar-powered system because it has the capacity to address nonlinear relationships and interactions between two or more input features. The model builds a pool of decision trees and averages their outcomes in order to enhance the accuracy and minimize over fitting, thus it can be used in complicated regression problems. Mean Absolute Error (MAE), the root mean squared error (RMSE), and the coefficient of determination(R 2) were used to evaluate the RF model. Optimalized Random Forest model attained MAE = 5.826 kg, RMSE = 7.974 kg, and R 2 = 0.9497, which means that it is a very good predictor with a little higher error than SVR model. This shows that the model explains most of the variations in the production of hydrogen as well as being very precise. The scatter plot of actual and predicted hydrogen production value of the Random Forest model is shown in Figure 5-5. The predicted values are near to the ideal 45-degree reference line, which means that there is a close correspondence between the predicted and the actual output over the observed range.

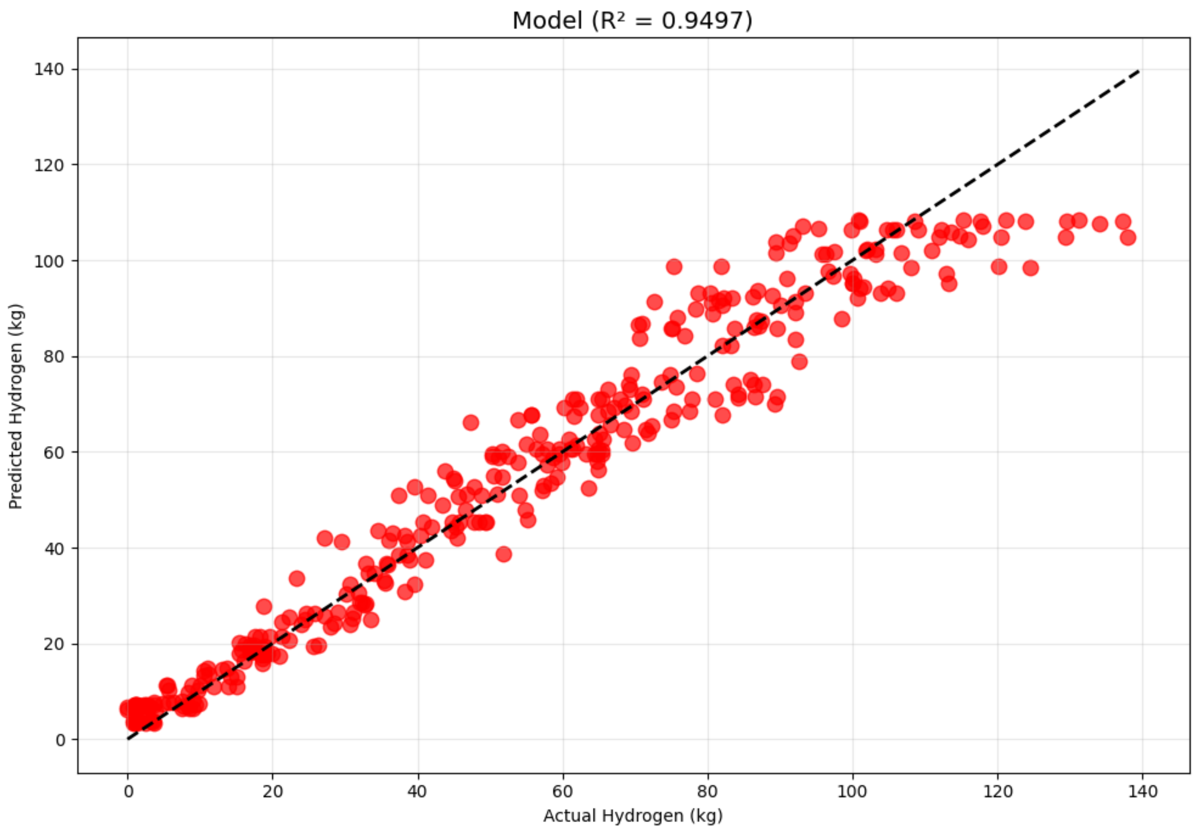

Figure 5-5Prediction vs. actual production

5.2.2 Decision Tree Regression Performance Evaluation for Hydrogen Production

To predict the production of hydrogen by the solar-powered system, the Decision Tree Regression (DT) was used because, like the other method, it is capable of modeling nonlinear relationships by segmenting the feature space into recursive partitions. The model produces simple decision rules to make approximations of the target variable and this makes it interpretable and easy to use in capturing complex pattern in the data. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination (R 2), were used to evaluate the DT model. The optimal model of the Decision Tree model was to use the optimized model which had an MAE of 5.295 kg, RMSE of 7.867 kg, and R 2 of 0.9511 meaning that the model is predictive and is capable of explaining a significant amount of the variation in hydrogen production. Figure 5-6 provides the scatter plot of the actual and the predicted values of hydrogen production using the Decision Tree model. The forecasts in most cases conform to the desired 45 degree referent line which signifies an acceptable consistency between forecasted and measured outputs throughout the measurement interval. Altogether, the Decision Tree model has good predictions of hydrogen production, which can model the nonlinear relationships in the dataset.

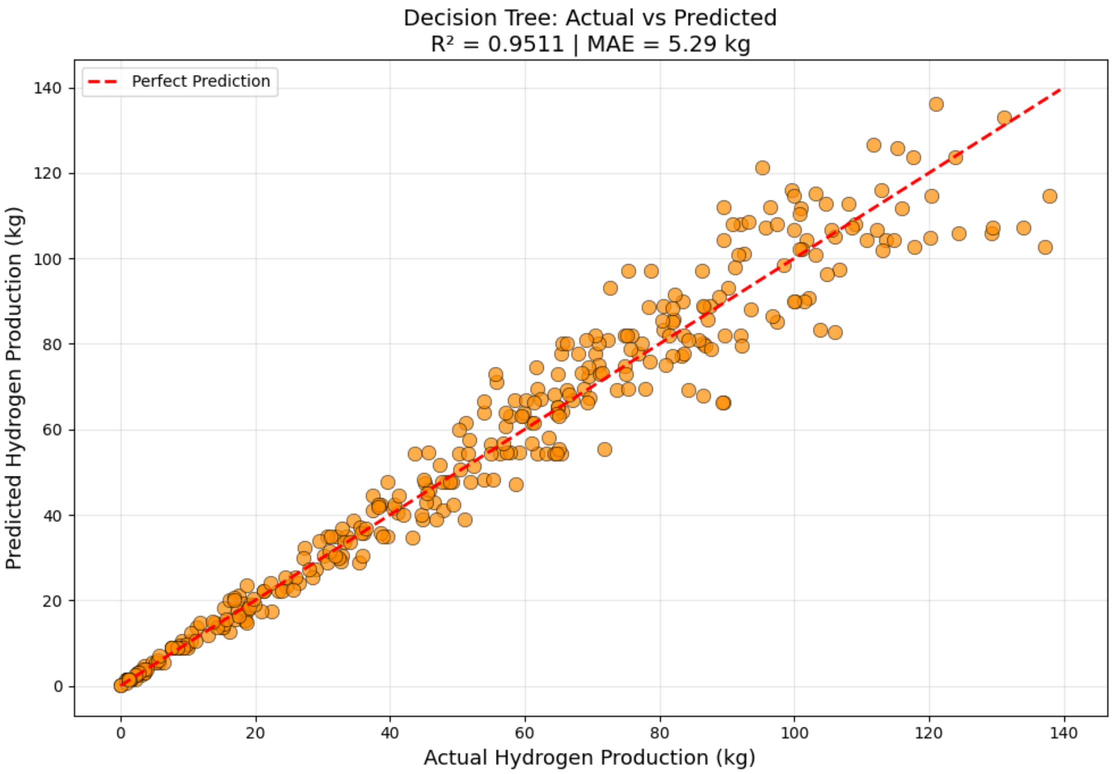

Figure 5-6 Prediction vs. actual production

5.2.3 GRU Neural Network Performance Evaluation for Hydrogen Production

A Gated Recurrent Unit (GRU) neural network was utilized to anticipate hydrogen production of the solar-powered system since it has the capacity to capture a time dependency in sequence-based data. GRU networks employ gating mechanisms to store pertinent information through time whilst alleviating the vanishing gradients, therefore appropriate in prediction tasks that vary with time. GRU model was tested based on Mean Absolute Error (MAE), root mean square error (RMSE), and the coefficient of determination (R 2) as well. The optimized GRU model gave an MAE of 9.273 kg, RMSE of 13.190 kg, and R 2 of 0.8378, which showed worse predictive performance than the other regression models. The scatter diagram 5-7 shows the real and the forecasted values of hydrogen production in the GRU model. Although the predictions reflect overall patterns, there is a higher degree of deviation around the desired 45 degree reference line, specifically at higher levels of hydrogen production per unit of time, which indicates the relatively low accuracy of the model. By and large, the GRU model shows average predictive power, as it successfully captures certain nonlinear and time-dependent trends in the data but is not the best compared to the SVR, Random Forest, and Decision Tree models.

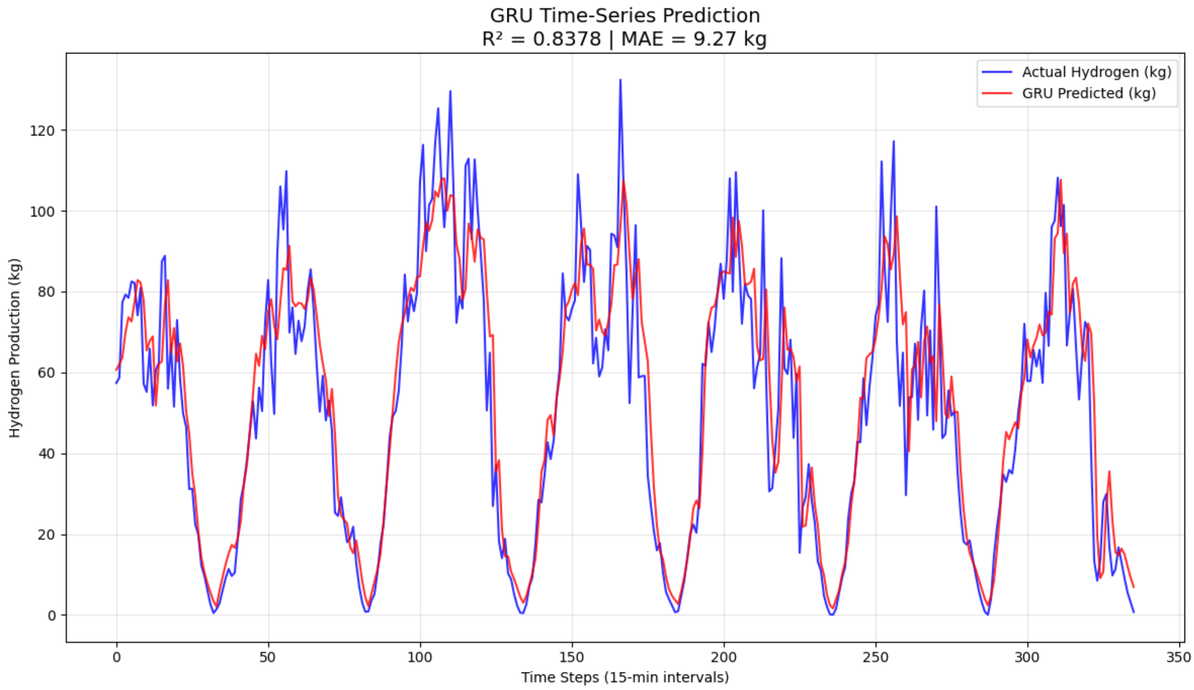

Figure 5-7Prediction vs. actual production

2.5.4 Support Vector Regression for Hydrogen Production

Out of the considered models, Support Vector Regression (SVR), Random Forest, Decision Tree, and GRU, the model that proved to be the most accurate predictor of hydrogen production through the solar-powered system is the SVR model. The SVR model had a Mean Absolute Error (MAE) of 4.713 kg, a Root Mean Squared Error (RMSE) of 6.498 kg and R 2 of 0.96476 implying that it predicts hydrogen production variance of approximately 96.48. The prediction error accounts for only 8.55 percent of the average output, with an overall production of 55.15 kg in 15 minutes; this shows that it is highly precise and reliable. Fig. 5-8 is a scatter chart of the actual and predicted values of hydrogen production with the SVR model with the predictions being closely clustered around the optimal 45-degree reference line indicating a great deal of agreement over the operating range. These findings prove that SVR is a better model compared to the rest in terms of addressing the nonlinear relationships between the input characteristics of the system such as module temperature, ambient temperature, solar irradiation and AC power and hydrogen output and therefore it is the best model to be used in this study.

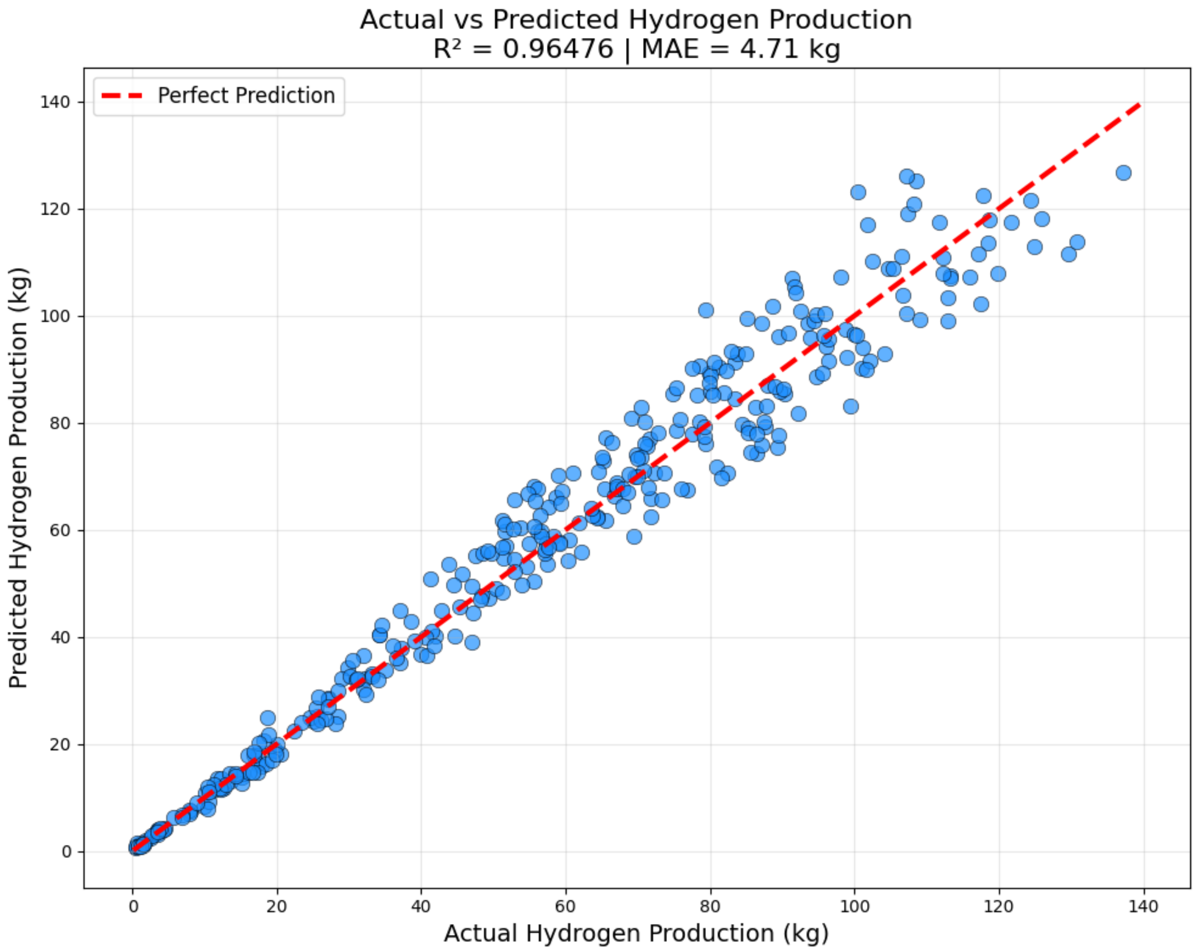

Figure 5-8 Prediction vs. actual production

Four models, including, SVR, Random Forest, Decision Tree, and GRU, were tested in order to predict hydrogen production. The SVR model had the lowest Mean Absolute Error (MAE = 4.713 kg) and Root Mean Squared Error (RMSE = 6.498 kg) and the highest R 2 of 0.9648, as indicated in the table. These findings show that SVR offers the best predictions and represents the largest amount of variance in hydrogen production, in case of the other models. Thus, it can be stated that SVR is the most effective model in the research, as stated in Table

Table 5-2Comparison table of ML results

Model	MAE (kg)	RMSE (kg)	R^2
SVR	**4.713**	**6.498**	**0.9648**
Random Forest	5.826	7.974	0.9497
Decision Tree	5.295	7.867	0.9511
GRU	9.273	13.190	0.8378

5.3Discussion

Table 5.2 provides the comparisons of hydrogen production prediction of four machine learning models, Support Vector Regression (SVR), Random Forest, Decision Tree, and Gated Recurrent Unit (GRU). The four models have been tested on the basis of performance evaluation parameters of MAE, RMSE and R 2. Each of these metrics will give an idea about the predictive performance and the capability of the two models to explain the variability in hydrogen production. The SVR model is the most effective of all the models. The lowest MAE (4.713 kg) and RMSE (6.498 kg) are obtained, which means that the average error in prediction is not very big and the resulting curves are not as sensitive to big deviations. In addition, the R 2 of SVR is 0.9648 meaning that the regression model accounts some 96.48 percent of the variability in hydrogen production. This positive result shows that SVR is useful in the process of modelling the non-linear relationship between the input features and the hydrogen output.

Random Forest and Decision Tree models are also quite good and have R 2 values of 0.9497 and 0.9511, respectively [31]. They, though, are more MAE and RMSE than SVR, which shows that they do not give the correct forecasts. Despite the fact that the Random Forest techniques are known to be advantageous due to the presence of ensemble learning and the superior performance of generalization in comparison with a single tree, it is worse than the SVR, possibly due to the complexity of the model or its vulnerability towards features that are varying in nature. The Decision Tree model is interpretable but fails at dealing with nonlinear patterns which are more complex and hence its relatively higher error rates. On the other hand, the worst performance is registered by the GRU model whose MAE (9.273 kg) and RMSE (13.190 kg), and R 2 (0.8378) are the highest. This suggests that the GRU architecture might not be the best with the given data, perhaps because it does not have enough depth in terms of time, is not trained on the most appropriate data, or lacks the strength to follow sequential dependencies within the hydrogen production process.[32].

Generally, the results indicate that the most effective model in this paper is SVR. Its efficiency in all measures of error and variance elucidation is a witness to its strength and reliability in the prediction of hydrogen production. Therefore, the best model that should be analyzed and implemented is SVR and it is summarized in Table5-2

Chapter 6
Conclusion and Future Work

6.1 Conclusion

The increased need of the world to clean and sustainable energy should make it incorporate innovative technologies to make maximum use of renewable energy. This paper demonstrates the possibility to produce surplus solar and wind energy into hydrogen to have a universal, and storable system of energy to deal with the intermittent problem of traditional renewable systems. Using the excess electricity produced at times of large wind generation due to solar or nocturnal winds, hydrogen energy does not only reduce energy wastage but also offer an avenue through which transportation and industrial sectors can be decarbonized.

The results highlight that the integration of renewable energy production with hydrogen production can increase the efficiency of the system as a whole, decrease the reliance on fossil fuels, and achieve energy security. Moreover, adoption of such form of hybrid systems could be utilized to drive the establishment of a robust hydrogen economy that will drive technological innovation and sustainable industrial growth. Such future studies must also aim at maximising the efficiency of the electrolysis process, incorporation of smart energy control mechanisms, and techno-economic viability of large-scale introduction of the renewable hydrogen to make sure that the renewable hydrogen becomes one of the viable building blocks of the low-carbon energy infrastructure.

6.2 Future Work

The cost of renewable hydrogen technology focuses on five factors. The optimization of electrolyzers would open the fast bright face of electrolyzers, especially under the variable situation of large scale solar and wind energy. Unused renewable energy is immediately transformed into hydrogen through new energy management schemes and this eliminates energy wastage. Other than that, the techno-economic analysis is also the main measurement device of costs, environmental implications, and the general scalability of these systems.

Hybrid storage solutions have also been taken into account to have a more flexible and resilient energy infrastructure combining hydrogen and battery storage [33]. Lastly, the research on the policies and incentives is invaluable towards identifying the policies and subsidies that to the best of its ability, contribute to the adoption of renewable hydrogen technologies.

References

[1] E. T. Sayed, A. G. Olabi, A. H. Alami, A. Radwan, A. Mdallal, A. Rezk, and M. A. Abdelkareem, "Renewable energy and energy storage systems," *Energies*, vol. 16, no. 3, art. no. 1415, Feb. 2023.

[2] A. G. Olabi and M. A. Abdelkareem, "Renewable energy and climate change," *Renewable and Sustainable Energy Reviews*, vol. 158, art. no. 112111, Apr. 2022.

[3] J. Ji, H. Liu, J. Guo, L. Jiang, and Z. Fan, "Multi-objective optimization study of regional integrated energy systems coupled with renewable energy, energy storage, and inter-station energy sharing," *Renewable Energy*, vol. 225, May 2024, art. no. 120328.

[4] A. Z. Arsad, M. A. Hannan, A. Q. Al-Shetwi, M. Mansur, K. M. Muttaqi, Z. Y. Dong, and F. Blaabjerg, "Hydrogen energy storage integrated hybrid renewable energy systems: A review analysis for future research directions," *Int. J. Hydrogen Energy*, vol. 47, no. 39, pp. 17285–17312, 5 May 2022.

[5] G. H. Guo and J. Guo, "Hydrogen gas harvest from the air – how did we turn an idea into reality?" in *Renewable Energy Innovations*, Berlin, Germany: Springer Nature, 2022, pp. 89–104.

[6] S. G. Simões, J. Catarino, A. Picado, T. F. Lopes, S. di Berardino, F. Amorim, F. Girio, C. M. Rangel, and T. Ponce de Leão, "Water availability and water usage solutions for electrolysis in hydrogen production," *J. Cleaner Prod.*, vol. 315, 15 September 2021, art. no. 128124.

[7] J. C. Bai, "Water electrolysis based on renewable energy for hydrogen production," *Chin. J. Catal.*, vol. 39, no. 3, pp. 390–394, Mar. 2018.

[8] S. B. Mizeraczyk and J. Mizeraczyk, "Overview of the hydrogen production by plasma-driven solution electrolysis," *Energies*, vol. 15, no. 22, p. 7508, Nov. 2022.

[9] B. Mahesh, "Machine learning algorithms – a review," *Int. J. Sci. Res. (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan. 2020.

[10] A. J. Bard and L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, 2nd ed. New York, NY, USA: Wiley, 2001, pp. 1–43.

[11] N. Pearsall, "Performance assessment of photovoltaic systems," in *The Performance of Photovoltaic (PV) Systems: Modelling, Measurement and Assessment*, Cambridge, UK: Woodhead Publishing, 2017, pp. 123–150.

[12] A. Javaid, U. J. Muhammad, M. S. Ejaz, Y. A. Khan, and A. W. Khan, "Forecasting hydrogen production from wind energy in a suburban environment using machine learning," *Int. J. Hydrogen Energy*, vol. 47, no. 95, pp. 40345–40356, Nov. 2022.

[13] Q. Yang et al., "Machine learning assisted prediction for hydrogen production of advanced photovoltaic technologies," *DeCarbon*, vol. 4, p. 100050, 2024.

[14] R. Ahmed, S. A. Shehab, O. M. Elzeki, A. Darwish, and A. E. Hassanein, "An explainable AI for

green hydrogen production: A deep learning regression model," *Int. J. Hydrogen Energy*, vol. 83, pp. 1226–1242, Aug. 2024.

[15] R. Zhang, X. Xu, Y. Zhang, and Y. Dong, "Analysis and forecast of the substitution potential of China's wind power-hydrogen production for fossil fuel hydrogen production," *J. Clean. Prod.*, vol. 422, p. 138410, Aug. 2023.

[16] M. S. Haider, S. I. Ahmad, and S. A. Haider, "Forecasting hydrogen production potential in Islamabad from solar energy using water electrolysis," *Int. J. Hydrogen Energy*, vol. 46, no. 2, pp. 1671–1681, Jan. 2021.

[17] H. Ishaq and I. Dincer, "Evaluation of a wind energy-based system for cogeneration of hydrogen and methanol production," *Int. J. Hydrogen Energy*, vol. 45, no. 32, pp. 15869–15877, Jun. 2020.

[18] H. Hassania and F. Saidi, "Feasibility of a standalone photovoltaic/Battery system with hydrogen production," *J. Energy Storage*, vol. 31, p. 101644, Oct. 2020.

[19] IEA, *World Energy Outlook 2023*, International Energy Agency, Paris, France, 2023.

[20] Z. Mei, J. Ding, L. Chen, T. Pi, and Z. Mei, "Hybrid multi-domain analytical and data-driven modelling for feed systems in machine tools," Symmetry, vol. 11, no. 9, Art. no. 1156, 2019, doi: 10.3390/sym11091156.

[21] M. Ball and M. Wietschel, *The Hydrogen Economy: Opportunities and Challenges*, Cambridge University Press, Cambridge, UK, 2009.

[22] UQSA — An R-Package for Uncertainty Quantification and Sensitivity Analysis for Biochemical Reaction Network Models, scientific figure, ResearchGate.[23] D. Stolten and B. Emonts, *Hydrogen Science and Engineering: Materials, Processes, Systems and Technology*, Wiley-VCH, Weinheim, Germany, 2016.

[24] K. Zeng and D. Zhang, "Recent progress in alkaline water electrolysis for hydrogen production and applications," *Prog. Energy Combust. Sci.*, vol. 36, no. 3, pp. 307–326, 2010.

[25] T. Markvart and L. Castaner, *Solar Cells: Materials, Manufacture and Operation*, Elsevier, Oxford, UK, 2005.

[26] A. F. Mirza, M. Mansoor, Q. Ling, B. Yin, and M. Y. Javed, "A salp-swarm optimization-based MPPT technique for harvesting maximum energy from PV systems under partial shading conditions," *Energy Conversion and Management*, vol. 209, Art. no. 112625, 2020, doi: 10.1016/j.enconman.2020.112625.

[27] G. B. Huang et al., "Forecasting solar power generation using machine learning techniques," *Energy Convers. Manage.*, vol. 148, pp. 131–141, 2017.

[28] S. Voyant et al., "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[29] M. Carmo, D. L. Fritz, J. Mergel, and D. Stolten, "A comprehensive review on PEM water electrolysis," *Int. J. Hydrogen Energy*, vol. 38, no. 12, pp. 4901–4934, 2013.

[30] R. Manabe et al., “Dynamic modelling of PEM electrolyzers under fluctuating power input,” *Appl. Energy*, vol. 226, pp. 993–1004, 2018.

[31] M. Ahmad et al., “Machine learning-based wind speed forecasting for green hydrogen production,” *Energy*, vol. 239, 2022, art. no. 121921.

[32] A. Khan et al., “Machine learning assisted photovoltaic optimization for green hydrogen production,” *Int. J. Hydrogen Energy*, vol. 47, no. 18, pp. 10234–10248, 2022.

[33] A. Mellit et al., “Smart photovoltaic–battery–hydrogen systems using artificial intelligence,” *Renewable and Sustainable Energy Reviews*, vol. 135, 2021, art. no. 11012.

[34] A. Siddiqa, N. Rana, W. Z. Khan, F. Jeribi, and A. Tahir, “Explaining solar forecasts with generative AI: A two-stage framework combining transformers and LLMs,” *PLOS ONE*, vol. 20, no. 9, e0331516, Sep. 2025.

[35] D. Rasch and O. Masata, “Methods of variance component estimation,” *Czech Journal of Animal Science*, vol. 51, no. 6, pp. 253–260, 2006.

[36] M. Mahmudi and A. R. Hidayat, “The Effect of Ensemble Averaging Method on Rainfall Forecasting in Jakarta Using ARIMA and ARIMAX,” *Mathline: Jurnal Matematika dan Pendidikan Matematika*, vol. 9, no. 1, pp. 35–46, 2024.

[37] H. Gupta, O. Sharma, D. Bhardwaj, and J. Yadav, “EEG Signal Based Multi-Class Emotion Recognition Using Hybrid 1D-CNN and GRU,” *Research Square*, 2023.

[38] M. Mohamed and K. Alosman, “A Comparative Study of Deep Learning Approaches for Arabic Language Processing,” *Jordan Journal of Electrical Engineering*, vol. 11, no. 2, pp. 145–156, 2025.

[39] M. H. Ghasemi, M. R. H. Yazdi, and F. M. Moosavian, “Support Vector Regression Model for Prediction of Groundwater Level Fluctuations,” *Applied Water Science*, vol. 13, no. 3, pp. 1–15, 2023.