Forecasting Hydrogen Production from Wind Energy in a Suburban Environment Using Machine Learning

Ali Javaid 1,2, Umer Javaid 1,2, Muhammad Sajid 1,2,*, Muhammad Rashid 3, Emad Uddin 1, Yasar Ayaz 1 and Adeel Waqas 4

1 School of Mechanical and Manufacturing Engineering (SMME), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
2 Artificial Intelligence for Mechanical Systems (AIMS) Laboratory, School of Interdisciplinary Engineering and Sciences (SINES), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
3 Department of Computer Science, National University of Technology (NUTECH), Islamabad 44000, Pakistan
4 U.S.-Pakistan Center for Advanced Studies in Energy (USPCAS-E), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
* Correspondence: m.sajid@smme.nust.edu.pk; Tel.: +92-51-886-6065

Abstract: The environment is seriously threatened by the rising energy demand and the use of conventional energy sources. Renewable energy sources including hydro, solar, and wind have been the focus of extensive research due to the proliferation of energy demands and technological advancement. Wind energy is mostly harvested in coastal areas, and little work has been done on energy extraction from winds in a suburban environment. The fickle behavior of wind makes it a less attractive renewable energy source. However, an energy storage method may be added to store harvested wind energy. The purpose of this study is to evaluate the feasibility of extracting wind energy in terms of hydrogen energy in a suburban environment incorporating artificial intelligence techniques. To this end, a site was selected latitude 33.64° N, longitude 72.98° N, and elevation 500 m above mean sea level in proximity to hills. One year of wind data consisting of wind speed, wind direction, and wind gust was collected at 10 min intervals. Subsequently, long short-term memory (LSTM), support vector regression (SVR), and linear regression models were trained on the empirically collected data to estimate daily hydrogen production. The results reveal that the overall prediction performance of LSTM was best compared to that of SVR and linear regression models. Furthermore, we found that an average of 6.76 kg/day of hydrogen can be produced by a 1.5 MW wind turbine with the help of an artificial intelligence method (LSTM) that is well suited for time-series data to classify, process, and predict.

Keywords: renewable energy; LSTM; forecast; artificial intelligence; machine learning; Islamabad

1. Introduction

Increasing energy demands resulting from rapid industrial digitization and urbanization of the world’s population coupled with the necessity of cleaner environments and emissions control has led to exponentially expanding interest in renewable and sustainable energy resources [1–3]. Hydro power and solar and wind energy constitute the main renewable sources of energy. Hydropower contributes 29%, and the cumulative contribution of solar and wind energy is 6% of the total energy consumption of Pakistan, as shown in Figure 1a based on data provided in [4]. These contributions the fulfilling of energy needs are expected to grow by 33% and 12%, respectively, by 2025, as indicated in Figure 1b based on data provided in [5]. Hydro power has a significant environmental impact on natural waterways, although dams are expensive to build and depend on natural reserves [6]. Whereas solar energy has a significant footprint throughout the day, during the night, another energy source is required to fulfil energy demands. Subsequently, environmental factors have a considerable impact on the efficiency of solar PV systems, including bifacial...
panels [7–9]. Wind energy provides a reasonable alternative; however, its integration into the energy grid is hindered by its intermittent nature [10].

![Energy mix of Pakistan (MW)](image_url)

**Figure 1.** Energy mix of Pakistan (MW) (a) in the year 2020 and (b) the expected mix in the year 2025.

In this study we propose a method to store wind energy in terms of hydrogen as a reasonable solution, as well as the use of artificial intelligence/machine learning (AI/ML) as a tool to assess the potential of the proposed solution. In conventional energy sources, energy storage occurs before electric power generation in the form of the primary resources of coal, gas, uranium, or oil. Electricity is generated according to demand by thermal power plants. However, the production of renewable energy from wind is detached from energy demand, and it is necessary to store energy after the wind turbines harvest it. A theme-based literature review is summarized in Table 1.

**Table 1. Theme-based literature review.**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Subtheme</th>
<th>Research Work and Findings</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Analytical method to produce hydrogen</td>
<td>An engineering equation solver (EES) was utilized to determine the overall energy and exergy efficiencies of the hydrogen production process, and the particle swarm optimization technique was used to optimize the system.</td>
<td>[11,12]</td>
</tr>
<tr>
<td>2</td>
<td>Biochemical hydrogen production</td>
<td>Biochemical methods play an important role in hydrogen production, in addition to being environmentally friendly. BEAMR was found to be an efficient method for hydrogen production. Moreover, the rate of hydrogen production can be increased by increasing the activity of hydrogen production bacteria.</td>
<td>[13–15]</td>
</tr>
<tr>
<td>3</td>
<td>Electrolyzer for the production of hydrogen</td>
<td>Proton-exchange membrane (PEM) electrolyzers were incorporated, for hydrogen production, and multiple electrolyzers were suggested. Electrolyzer were compared, revealing that an electrolyzer with a membrane is most efficient. Hydrogen production of up to 1.2 kg/h can be achieved using a PEM electrolyzer, but with Cr-C coated SS304 bipolar plates in chemical solution, a 1 MW system was able to produce 6 m³/h. Methanol was also introduced in PEM for hydrogen production.</td>
<td>[16–20]</td>
</tr>
<tr>
<td>4</td>
<td>Hydrogen production</td>
<td>Multiple methodologies were presented, and mostly renewable sources were utilized to produce hydrogen as a product with some by products, such as fresh water and electricity. Moreover, the prospective and attractiveness of hydrogen production was elaborated.</td>
<td>[21–27]</td>
</tr>
<tr>
<td>5</td>
<td>Solar energy for hydrogen production</td>
<td>Machine learning was applied to forecast hydrogen production, yielding the Prophet model with a daily average of 93.3 × 103 kg/km². Furthermore, a solar-driven, steam-autothermal hybrid reforming system (SAHRS) was proposed to capture the carbon emissions generated during hydrogen extraction. It was also concluded that solar energy had a considerable impact for superheating and steam generation to feed gaseous reactants.</td>
<td>[28–31]</td>
</tr>
</tbody>
</table>
Researchers are continuously working on improved methods for clean, green energy sources. The focus of this study is a novel approach to hydrogen harvesting using artificial intelligence for its efficient production. Using wind (speed, direction, gust, etc.) as inputs machine learning algorithms forecast hydrogen production in Islamabad (Pakistan's Capital) as their output. Furthermore, the proposed method saves time and implementation costs by pre-estimation of clean green energy through a mathematical model based on sample data. An added advantage of implementing machine learning is that with the help of historical wind data, computer systems can perform special tasks to identify the pattern of wind speed and automatically predict the next stage, automatically learning and improving on past experiences. Among others, long short-term memory (LSTM), support vector regression (SVR), and linear regression architectures are used in this particular study, in consideration of the continuous data type.

2. Methodology

Prior to forecasting, one year of data was acquired to ensure that the seasonal variation was captured for a suburban environment. The primary input for the machine learning models was a series of data sets that were collected over the course of year to capture seasonal variation. The data sets include values for wind gust, wind direction, and wind speed that were recorded in 10 min intervals for the optimal prediction outcomes. The data sets of wind speed, wind direction, and wind gust acquired with the help of an anemometer were run through machine learning algorithms to train the models. Among all the trained
models, the best-predicting model was utilized for estimation of hydrogen production implementing PEM. A comparison was drawn with estimated hydrogen production from the measured wind speed data acquired from the instrument and the predicted hydrogen production potential of AI-predicted wind speed for the same time span. The obtained results are discussed in subsequent sections.

2.1. Apparatus

The apparatus used to quantify the speed of wind was an NRG 40H anemometer [55], which is suitable for an electrically noisy environment and compatible with instruments that require a square wave signal. It is also used in small wind turbines as a control sensor, whereas in high-speed winds, it triggers the PV tracker to stow. A high-level square-wave voltage signal is generated that is compatible with the frequency of the wind speed. A wide range of industrial instruments is compatible with the signal. The instrument specifications are as follow: 3-cup anemometer sensor with a range of 1 m/s to 96 m/s (2.2 mph to 215 mph), and output signal range of 0 Hz to 125 Hz, and a threshold speed of 0.78 m/s (1.75 mph). The operating conditions were as follows: supply voltage, 5 V to 26 V DC; supply current, 9 mA maximum; operating temperature and humidity ranges, \(-55 \degree C \) to \(60 \degree C \) (\(-67 \text{ F} \) to 150 F) and 0 to 100% RH, respectively. The swept diameter of the rotor is 190 mm (7.5 inches).


The long short-term memory (LSTM) model was used in deep learning as an artificial recurrent neural network. LSTM is distinct from the other models, as it includes feedback connections. LSTM has a wide range of applications, such as handwriting recognition, network traffic detection (anomaly), intrusion detection, etc. Essentially, gates of an LSTM unit comprise a cell, input gate, forget gate, and output gate. The cell stores the values for a certain period of time, and the gates process the information into and out of the cell. The network serves best in categorizing, processing, and forecasting based on time series data. A detailed study was presented by Young on LSTM cells and architecture [56]. The hyperparameters, which were set before the training of the model are hidden neuron: 300, layers: 2, and Learning rate: 0.01.

Support vector regression (SVR) is another technique used to predict wind speed for hydrogen production. In SVM, the rate of error is minimized, whereas in the case of SVR, the error is fixed within a certain limit. This enables definition the extent to which an error is acceptable in our model to fit the data. The maximum error we set is called epsilon, and it can be adjusted to obtain the desired accuracy. Another method adopted for prediction of hydrogen production is linear regression, which is used to predict relationships between two variables. The technique involves the observation of data by fitting a linear equation. All the methodologies adopted herein are used to compare hydrogen production forecasting methods. The methods mentioned above can provide an overview of the extent to which wind energy can be harvested in a particular area, resulting in hydrogen production. Artificial intelligence enables the use of the information predicted by the models, such that if a reasonable amount of hydrogen production is predicted for a particular site, only then can the installation of turbines be executed. This methodology saves time and cost associated with the production of green energy.

A proton-exchange membrane (PEM) electrolyzer utilizes the electrical energy generated by wind turbines and produces hydrogen. The recorded data sets from the sensor (data measured with the help of an NRG 40H anemometer) and predicted data sets from AI models were compared for validation of the results, which are presented in Section 3. For ease of understanding and visualization, MATLAB was used to generate graphs. The whole mechanism can be separated into three stages, i.e., the wind turbine, the inverter (AC/DC), and the PEM electrolyzer (Figure 2). Figure 3 presents the workflow diagram displaying the process involved in the prediction process of wind speed data sets. The data sets were acquired by an NRG40H anemometer and fed into the training models (LSTM,
SVR, and LR). The generation of hydrogen was then quantified using the predicted data that was acquired through the machine learning approach.

![Figure 2. Schematic diagram of the adopted approach.](image)

![Figure 3. Workflow diagram for forecasting hydrogen production.](image)

### 2.3. Harvesting Hydrogen from Wind Power

To obtain the results, some of the parameters were fixed; for example the efficiency of electrolyzer was set as 75% [57], and the efficiency of the wind turbine ranged from 35% to 45%, as in [47]. In this study, the turbine efficiency was set at 40%, and the rotor diameter was set at 77 m [58]. The rotor diameter plays a significant role in increasing or decreasing wind energy. As the diameter increases, the swept area also increases, resulting and an upsurge in wind energy and vice versa. Hence, a direct relationship exists between
rotor diameter and wind energy. The lower heating value of hydrogen was designated as 33.3 kWh/kg, and the density of air was set to 1.225 kg/m$^3$.

A wind turbine is used to convert mechanical power into electrical power. The electrical output power can be estimated according to [47] and [59]:

$$ P_{wt} = \frac{1}{2} V^3 \eta_{wt} \rho_{air} A_{wt} C_{p,wt} $$  \hspace{1cm} (1)

where $V$ is the speed of wind measured in m/s; $\eta_{wt}$ is the efficiency of the wind turbine, assumed to be 40%; $\rho_{air}$ is the density of air, equal to 1.225 kg/m$^3$; $A_{wt}$ is the swept area; and $C_{p,wt}$ is wind turbine power coefficient. The following relation can be used to compute the net input energy.

$$ P_{input} = \frac{1}{2} \rho_{air} A_{wt} V^3 $$  \hspace{1cm} (2)

The power output by the generator of the wind turbine is further processed through an AC/DC converter and utilized by the proton-exchange membrane electrolyzer (PEM). The PEM splits the water into its basic constituents, i.e., hydrogen and oxygen, through electrolysis. The electricity used by the PEM is obtained from the wind turbines. The rate of hydrogen production gained by the PEM electrolyzer can be computed with the help of the following relation [60].

$$ \dot{N}_{H_2} = \frac{J_{el}}{2F} $$  \hspace{1cm} (3)

where $J_{el}$ is the density of the current, $F$ is Faraday’s constant, and $\dot{N}_{H_2}$ represents the hydrogen mole flow. The amount of hydrogen mass produced from the input of wind energy in kg/h is given by [57,58].

$$ M_{H_2} = \frac{\eta_{el} P_{wt}}{LHV_{H_2}} $$  \hspace{1cm} (4)

where $M_{H_2}$ represents the produced mass of hydrogen, $\eta_{el}$ indicates the efficiency of the electrolysis process, $P_{wt}$ is the wind power input to the electrolyzer, and by $LHV_{H_2}$ represents the lower heating value for hydrogen. Another relation used to estimate wind turbine output energy is [58].

$$ P_{wt} = E_{out} $$  \hspace{1cm} (5)

$$ E_{out} = C_F P r T $$  \hspace{1cm} (6)

Here, $P_r$ is defined as the wind-turbine-rated power, and $C_F$ is the capacity factor [58]. The cost of production of wind energy can also be estimated by the capacity factor.

$$ C_F = \frac{P_{out}}{P_r} $$  \hspace{1cm} (7)

The overall energy and exergy efficiencies of wind-power-based hydrogen production can be calculated as follows [61].

$$ \eta_{wind} = \frac{m_{H_2} LHV_{H_2}}{P_{input}} $$  \hspace{1cm} (8)

$$ \Psi_{wind} = \frac{m_{H_2} \text{ex}_{H_2}}{P_{input}} $$  \hspace{1cm} (9)
3. Results and Discussions

The aim of this study was to utilize a clean, green energy source and store it in an environmentally friendly, cost-effective manner. The prediction of wind speed leads us to estimate the hydrogen production for the city of Islamabad. The technique adapted in this study is flexible and can be implemented for any region in the world. Predicted wind speed estimates the hydrogen production using three machine learning algorithms. One added advantage of this research approach is that we can forecast wind direction and install fixed or unidirectional wind turbines to optimize generation. The graphs for absolute errors and standard deviation of wind speed and wind direction are shown below.

K-fold cross validation was applied to algorithms used in this research. The results are discussed below, with 10 folds taken for each case. Figure 4 shows the wind speed variation in K folds compared with the mean absolute error and standard deviation among all the machine learning approaches applied in this study. Similarly, Figure 5 shows the wind direction variation in K folds comparing the mean absolute error and standard deviation of the wind direction.

Results show in Figures 4 and 5 clearly indicate that of all the techniques applied in this study, LSTM has the lowest mean absolute error and standard deviation, with a consistent line, i.e., no fluctuations as compared to the other techniques. The performance of the machine learning methods was evaluated based on the mean absolute error and standard deviation. Among all the models, LSTM has the lowest mean absolute error of up to 0.7 m/s, whereas the other models reached up to 1.1 m/s with fluctuations.

Furthermore, based on the accuracy graph, LSTM was chosen for prediction of wind speed and wind direction. Figure 6 demonstrates the accuracy of the measured wind speed vs. predicted wind speed. The average accuracy was found to be 78.66%. Data for the preceding two days are taken into account for the prediction of the succeeding two days. The interval time is set to 10 min.

![Figure 4. Wind speed variation in K folds.](image-url)
Results show in Figures 4 and 5 clearly indicate that of all the techniques applied in this study, LSTM has the lowest mean absolute error and standard deviation, with a consistent line, i.e., no fluctuations as compared to the other techniques. The performance of the machine learning methods was evaluated based on the mean absolute error and standard deviation. Among all the models, LSTM has the lowest mean absolute error of up to 0.7 m/s, whereas the other models reached up to 1.1 m/s with fluctuations.

Figure 4. Wind speed variation in K folds.

Figure 5. Wind direction variation in K folds.

Furthermore, based on the accuracy graph, LSTM was chosen for prediction of wind speed and wind direction. Figure 6 demonstrates the accuracy of the measured wind speed vs. predicted wind speed. The average accuracy was found to be 78.66%. Data for the preceding two days are taken into account for the prediction of the succeeding two days. The interval time is set to 10 min.

The data sets of measured wind speed and predicted wind speed are plotted together for ease of understanding. The predicted wind speed conforms closely to the actual, measured wind speed, as shown in Figure 7. Furthermore, a bar chart was used to represent the average hydrogen produced in an hour per day in Figure 8. With the exception of a few bars caused by unusual wind gusts, the predicted values of average wind speed per day are roughly comparable to the measured values. Barring unusual wind gusts, there is a small margin between the predicted and observed values.

Figure 6. Accuracy of average wind speeds per day for the month of August.

Figure 7. The accuracy of measured wind speed vs. predicted wind speed.

Figure 8. Average hydrogen produced in an hour per day.
The relationship between the average hydrogen production per hour in one day vs. total hydrogen production per day appears to be closely linked, as indicated in Figure 9. These results were obtained in Islamabad; however, this forecasting method can be applied to an optimal location for maximum hydrogen production. By forecasting hydrogen production in such a manner, investments can be made at optimal locations to harvest more energy. Figure 10 shows the rate of hydrogen production by the predicted wind speed for the month of August. The same graph is expanded for improved visualization, as shown in Figure 11, which demonstrates the rate of hydrogen production on a daily basis. The result shows that the rate of production increases during daytime and decreases at night, and the rate of hydrogen reaches its maximum value around noon. The rate of production starts increasing at 1000 h and starts decreasing almost immediately after noon.
The rate of production starts increasing at 1000 h and starts decreasing almost immediately after noon. The present results may be helpful in predicting other renewable sources with the models used herein or with other models. For this purpose, considerable computational resources will be required for comparative studies to train multiple models incorporating different data set sampling frequencies, time periods, locations of data acquisition, etc. Furthermore, hydrogen storage methods can also be studied to harvest and store energy for later use.

Figure 9. Average hydrogen production and total hydrogen production.

Figure 10. Estimated rate of hydrogen production.

Figure 11. Estimated rate of Hydrogen production per day.

4. Conclusions

In this study, we emphasized the use of machine learning methods to forecast the hydrogen production by a PEM electrolyzer using wind energy from wind turbines. For this purpose, wind speed and wind direction data were collected with the help of an NRG 40 anemometer and an NRG 200 wind direction sensor, respectively, at latitude 33.64°N, longitude 72.98°N, and altitude 500 m above mean sea level. One year of data were K-fold cross-validated using three methods i.e., linear regression, SVR, and LSTM, which was used to forecast the hydrogen production. The output of the model was used to estimate wind energy production in the area under consideration and identify the minimum, maximum, and average energy capacity. Among all the models, LSTM had the lowest MAE and standard deviation wind speed, with values are consistent at around 0.72 m/s and 0.68 m/s, respectively. By implementing the LSTM model, we found that an average of 6.76 kg/day of hydrogen can be produced from a single wind turbine in a selected suburban environment. Moreover, a strong correlation is indicated between wind speed and solar energy, as shown in Figure 11. The wind speed reaches its maximum in the afternoon, resulting in increased hydrogen production, whereas a second production peak occurs around 0200 hrs. The LSTM model provides a satisfactory estimate of hydrogen production potential at the selected site; such forecasts are essential for project planners to effectively harness green energy sources.

Author Contributions:


Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.
The present results may be helpful in predicting other renewable sources with the models used herein or with other models. For this purpose, considerable computational resources will be required for comparative studies to train multiple models incorporating different data set sampling frequencies, time periods, locations of data acquisition, etc. Furthermore, hydrogen storage methods can also be studied to harvest and store energy for later use.

4. Conclusions

In this study, we emphasized the use of machine learning methods to forecast the hydrogen production by a PEM electrolyzer using wind energy from wind turbines. For this purpose, wind speed and wind direction data were collected with the help of an NRG 40 anemometer and an NRG 200 wind direction sensor, respectively, at latitude $33.64^\circ$ N, longitude $72.98^\circ$ N, and altitude 500 m above mean sea level. One year of data were K-fold cross-validated using three methods i.e., linear regression, SVR, and LSTM, which was used to forecast the hydrogen production. The output of the model was used to estimate wind energy production in the area under consideration and identify the minimum, maximum, and average energy capacity. Among all the models, LSTM had the lowest MAE and standard deviation wind speed, with values are consistent at around 0.72 m/s and 0.68 m/s, respectively. By implementing the LSTM model, we found that an average of 6.76 kg/day of hydrogen can be produced from a single wind turbine in a selected suburban environment. Moreover, a strong correlation is indicated between wind speed and solar energy, as shown in Figure 11. The wind speed reaches its maximum in the afternoon, resulting in increased hydrogen production, whereas a second production peak occurs around 0200 hrs. The LSTM model provides a satisfactory estimate of hydrogen production potential at the selected site; such forecasts are essential for project planners to effectively harness green energy sources.


Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{wt}$</td>
<td>Swept area</td>
</tr>
<tr>
<td>$C_F$</td>
<td>Capacity factor</td>
</tr>
<tr>
<td>$C_{p,wt}$</td>
<td>Power coefficient of wind turbine</td>
</tr>
<tr>
<td>$F$</td>
<td>Faraday's constant</td>
</tr>
<tr>
<td>$LHV_{H_2}$</td>
<td>Lower heating value</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long short-term memory</td>
</tr>
<tr>
<td>LR</td>
<td>Linear regression</td>
</tr>
<tr>
<td>$I_{el}$</td>
<td>Current density</td>
</tr>
<tr>
<td>$M_{H_2}$</td>
<td>Mass of hydrogen</td>
</tr>
<tr>
<td>$N_{H_2}$</td>
<td>Rate of hydrogen production</td>
</tr>
<tr>
<td>OTEC</td>
<td>Ocean thermal energy conversion</td>
</tr>
<tr>
<td>$P_r$</td>
<td>Rated power of the wind turbine</td>
</tr>
<tr>
<td>$P_{wt}$</td>
<td>Output power from wind turbine</td>
</tr>
<tr>
<td>PEM</td>
<td>Proton-exchange membrane</td>
</tr>
<tr>
<td>SVR</td>
<td>Support vector regression</td>
</tr>
<tr>
<td>$\eta_{el}$</td>
<td>Efficiency of the electrolysis process</td>
</tr>
</tbody>
</table>