Solar Irradiation Prediction Using Machine Learning Techniques: The Case of Saudi Arabia

Tayeb Brahimi
Energy and Technology Lab
College of Engineering
Effat University
Jeddah 22332, KSA
tbrahimi@effatuniversity.edu.sa

Ahmed Bensenouci
Electrical and Computer Engineering
College of Engineering
Effat University
Jeddah 22332, KSA
abensenouc@effatuniversity.edu.sa

Budoor Alwateed
Master Student, Master in Energy Engineering
College of Engineering
Effat University
Jeddah 22332, KSA
bualwated@effat.edu.sa

Fatima Alhebshi
Bachelor Student, Electrical and Computer Engineering
College of Engineering
Effat University
Jeddah 22332, KSA
faalhebshi@effat.edu.sa

Abstract

The Kingdom of Saudi Arabia recently set ambitious targets in its national transformation program and Vision 2030 to move away from oil dependence and redirect oil and gas exploration to other higher-value uses to supply 10% of its energy demand from renewable sources. The incorporation of solar energy into the grid becomes important due to its ever-increasing demand growth. This paper proposes to use machine learning techniques to predict daily GHI in some cities in Saudi Arabia. The paper compares the prediction of GHI using Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Random Forest, and Random Tree. The data used in this study uses attributes such as variable weather and solar irradiation data provided by KACARE. A comparison between actual and predicted GHI revealed that Random Forest performed better than the rest of the machine learning techniques with a root mean square error (RMSE) and a correlation coefficient of (45.560, 0.990) respectively compared to ANN (106.709, 0.984), SVM (203.79, 0.77), Random Tree (80.326, 0.968). The significance of the study relies on its ability to predict solar GHI for sustainable integration of PV systems the electrical grid and help operators manage the power generated more efficiently.

Keywords

Solar Energy, GHI, Artificial Intelligence, PV Systems, Machine Learning

1. Introduction

Despite all the progress made in the development of renewable and non-renewable energy, many areas around the world still lack electricity, in particular, in sub-Saharan Africa and rural areas. One of the best-decentralized and least-cost solutions is to use solar photovoltaic (PV) in an off-grid system. The Middle East and North Africa (MENA) region is considered as one of the best areas for solar radiation, in the world, as seen in Figure 1 (Vaisala, 2015; Zell et al., 2015). Sustainable Development Goals (SDGs), adopted by 193 nations in 2015, include, for the first time, a target to ensure access to affordable, reliable, and modern energy for all by 2030 (UN, 2015). In an analysis by IRENA (2018) on the investment opportunities in the GCC, it is reported that close to 60% of the GCC's surface is hit by the sun and has significant potential for solar PV power systems. The European Photovoltaic Industry Association (EPIA) and Greenpeace expect that PV could provide up to 12% of electricity demand in European countries by 2020 (Greenpeace, 2011). Developing just 1% of this technology could eventually result in 470 GW of solar PV capacity.

This paper attempts to use machine learning techniques to estimate the daily solar irradiation in some cities in Saudi Arabia. By using historical data, it is possible to build models capable of predicting and generating rules that can be translated into natural query language and provide a measure of the confidence of the classification based on its attributes. The data used in this study uses attributes such as variable weather and solar irradiation data provided by KACARE (2015) as part of the Renewable Resource Monitoring and Mapping (RRMM) Program. The paper compares the prediction of GHI using Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Random Forest, and Random Tree.

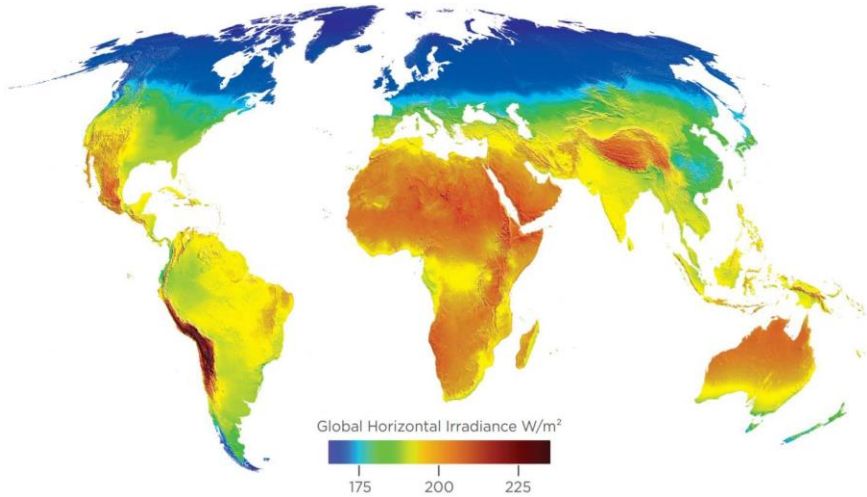

Figure 1. Worldwide global Solar map (Vaisala, 2015)

2. Vision 2030 and Saudi Renewable Energy Target

In Saudi Arabia, 15% of the domestic oil production is used to generate electricity, and another 50% is consumed by electric power plants (Lashin et al., 2015; IEA, 2016). According to the Electricity and Cogeneration Regulatory Authority, the residential sector consumes 50% of the generated power (KAPSARC, 2017). In the last ten years, the consumption per Capita in the Kingdom increased from 7,258 kWh/Capita to 9,167 kWh/Capita. In 2017, Saudi Arabia produced only 0.04% of electricity from non-fossil fuels using solar energy (KAPSARC, 2017). Recently, Saudi Arabia set an ambitious target in its national transformation program and vision 2030 to move from oil dependence and redirect the oil and gas exploration to other higher-value uses (KSA, 2017a; KSA 2017b). This goal is being achieved by setting an energy roadmap with the aim to supply 10% of its energy demand from renewable sources.

In its vision 2030 (KSA, 2017a), Saudi Arabia identified renewable energy as one of the pillars of economic diversification, away from oil, with an initial target set to 9.5 GWs of renewable energy by 2023 and 3.45 GW by 2020. The abundance of solar resource potential in KSA combined with the falling cost of associated technologies, including photovoltaic (PV) modules, represents the major factor behind the increased use of this source of energy. Recently, Saudi Arabia launched one of the Kingdom's largest projects to build $500 billion mega-city business and industrial zone "NEOM" with a high ambition to make this new area running on 100% renewable energy. As reported in National Renewable Energy Program (NREP, 2019), Saudi Arabia significantly increased its renewable energy targets and long-term visibility with 12 pre-developed projects to be tendered in 2019 with a capacity of 3.1 GW. The Renewable Energy Project Development Office (REPDO) of Saudi Arabia's Ministry of Energy, Industry and Mineral Resources (MEIM, 2019) announced a substantial increase in the renewable energy share to produce 40 GW of solar energy and 16 GW of wind power over the next decade as shown in Figure 2.

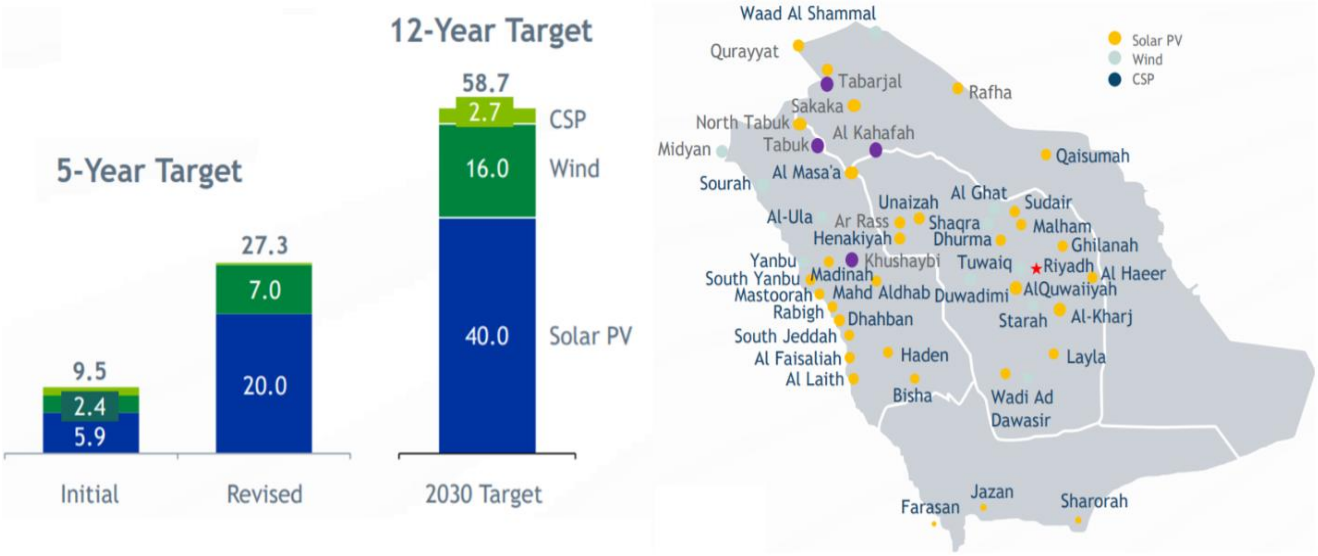

Figure 2. National Renewable Energy Program (NREP, 2019)

3. Related Work

The integration of solar energy into the electrical network would be more efficient if the fluctuation of the Global Horizontal Irradiation/Irradiance (GHI) is more reliable, and hence the PV energy output predicted more accurately. Although a solar PV system relies on many components such as inverter, charge controller, batteries, and panels, forecasting the solar irradiation represents the main step in ensuring and designing an efficient solar PV system since it depends on several parameters namely, PV system location and orientation, daytime, and sunshine period (Shehata et al., 2019; Sanyour et al., 2019). In recent years, many researches have been devoted to the study of solar energy management by using artificial intelligence techniques to maintain a sustainable integration of solar energy into the electricity grid (Mosavi, et al. 2019). The most popular forecasting methods reported in the literature are classified into four categories: physical methods, statistical methods, hybrid methods, and artificial intelligence methods (Alencar et al. 2017; Mosavi, et al. 2019). The core strategy in artificial intelligence methods is to build a relationship between the input and output data using some algorithm rather than using analytical methods. In statistical methods, the model looks for decreasing the difference between the predicted and immediate past value using auto-recursive mathematical model, while neural networks look for patterns in the input and output data over a long period of time (Mosavi, et al. 2019).

The Artificial Neural Network (ANN) model, Figure 3, is inspired by the human brain. It is composed of a number of connected layers and neurons to find relations between predicted and actual data. The Hidden layers with their corresponding weights are located using hidden sigmoid units (Alpaydin, 2014).

In the Support Vector Machine (SVM) algorithm, a supervised learning model, data used for classification and regression are analyzed. The algorithm solves linear and non-linear problems by separating the data into classes and creating a hyperline maximizing the margin between the two classes (Wu et al., 2008). SVM is considered a very powerful tool due to its high generalization capability, having a rather simple geometrical interpretation, a sparse solution, and the ability to process high dimensional data. SVM generated results are highly stable and reproducible. SVM finds its applications in a wide variety of fields and has successfully been applied in many real world problems such as pattern recognition, text and webpage classifications, weather prediction, and data mining.

The Random Forest (RF) algorithm combines Breiman's "bagging" idea and the random selection of features (Ayyadevara, 2018). It operates by constructing a multitude of decision trees at training stage and outputting the class that is the mode of the classes (classification) of the individual trees. It corrects for overfitting to their training set (models training data too well). Some advantages are: runs efficiently on large databases, highly efficient in the presence of missing data, can detect variable interactions, does not require data processing, etc.

Random Trees (RT) algorithm deals with both classification and regression problems (Fratello & Tagliaferri, 2019). A supervised Classifier is an ensemble-learning algorithm that generates many individual learners. RT takes

the input feature vector, classifies it with every tree in the forest, and outputs the class label that received the majority of "votes". RT is one of the most popular learning methods commonly used for data processing.

4. Solar Data Collection and Analysis

In the present study, data have been collected from the King Abdullah City for Atomic and Renewable Energy as part of the Renewable Resource Monitoring and Mapping (RRMM) Program (KACARE, 2017). The hourly data provided (May 2013 to July 2016) include different attributes such as air temperature, wind direction and speed, Global Horizontal Irradiance (GHI), relative humidity, and barometric pressure.
In designing solar power plants, the sun irradiation should be measured at the planned site. There are two types namely, PhotoVoltaic PV solar panel and Concentrated Solar Panel (CSP). CSP systems generate solar power using mirrors/lenses to concentrate a large area of sunlight onto a small area. Electricity is generated when the concentrated light is converted to heat, which drives a steam turbine connected to an electrical power generator. PV, on the other hand, absorbs sunlight as a source of energy to generate direct current electricity. The total amount of sun radiation (insolation) received from above by a horizontal surface is the Global Horizontal Irradiation (GHI). This value combines both Direct Normal Irradiation (DNI) and Diffuse Horizontal Irradiation (DHI) as

$$GHI = DHI + DNI \cdot cos(\theta) \tag{1}$$

where θ is the solar zenith angle defined as the angle between the zenith and the center of the sun. In fact, DNI is important for the CSP, whereas DHI is important for PV panels. The higher the insolation, the higher will the energy yield be. The later is affected by the wind speed and direction, air temperature, precipitation and soiling (sand storms)

Figure 3 shows the hourly distribution of GHI for the city of Jeddah, King Abdulaziz University Station (KAU, longitude 21.49604, latitude 39.24492) from June 2013 to July 2016. The graph shows the importance of the GHI intensity, where the maximum reaches more than 1000 Wh/m^2. This amount of solar radiation plays an important role in the design of PV systems. Figure 4 shows the average monthly temperature and maximum monthly temperatures for King Abdulaziz University Station. Increasing temperature increases the current, reduces the voltage, and reduces the power output. However, as the intensity of the sun increases, the output power generated by the PV system increases.

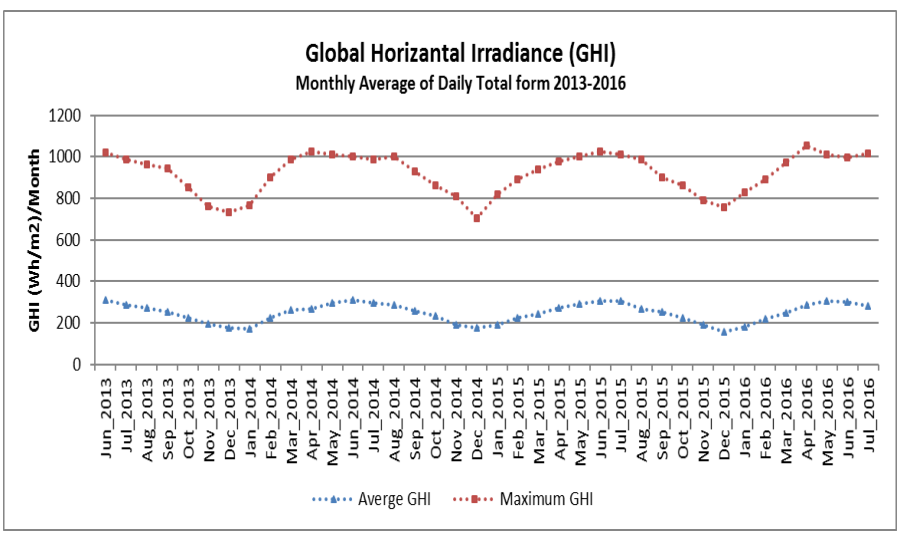

Figure 3. Monthly GHI distribution

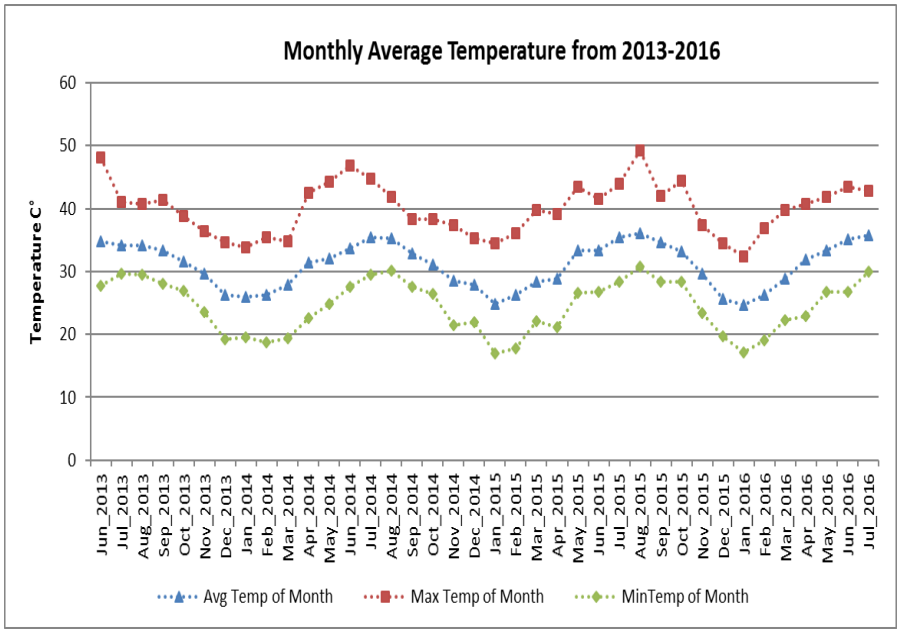

Figure 4. Monthly temperature distribution

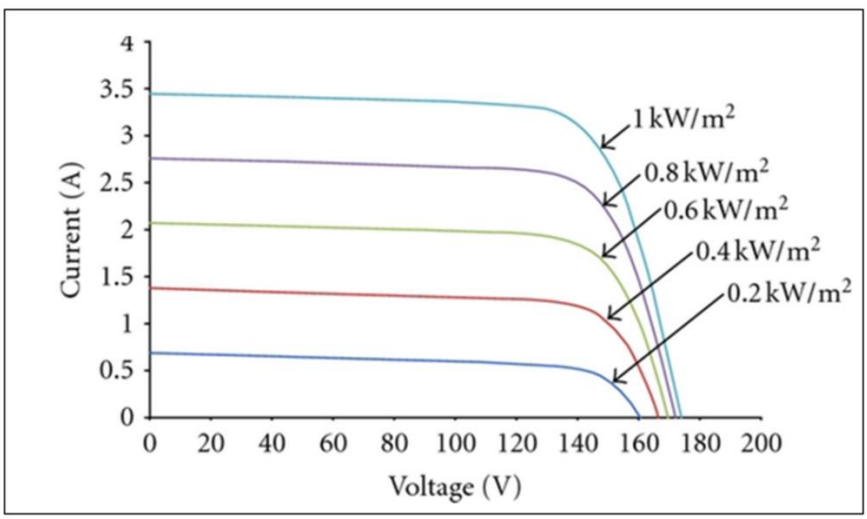

Figure 5.Effect of solar irradiation on the IV curve

5. Machine Learning Algorithms

Four machine learning Algorithms have been used to predict the GHI for the city of Jeddah, KAU station, namely, ANN, Random Forest, and Random Tree. Figure 6 shows an example of multi-layer neural networks formed by three layers, a passive input layer with seven nodes and three hidden-layers with 8, 4, and 3 neurons, respectively, and finally, an active output layer for the wind speed prediction. The performance of each proposed algorithm is measured by using the Root-Mean-Square-Error (RMSE) and the correlation coefficient (R). Values close to zero are appropriate for RMSE, while values close to one for R indicate a strong correlation. Assuming GHI_i^{act} is the wind speed observed, GHI_i^{pred} is the GHI predicted, the root mean square root is given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(GHI_i^{act} - GHI_i^{pred}\right)^2} \tag{2}$$

Using the average $\overline{\mathrm{GHI}^{\mathrm{act}}}$ and $\overline{\mathrm{GHI}^{\mathrm{pred}}}$ speeds the linear correlation coefficient can be rearranged as

$$R = \frac{\sum_{i=1}^{N}(\mathrm{GHI}_i^{\mathrm{act}} - \overline{\mathrm{GHI}^{\mathrm{act}}})(\mathrm{GHI}_i^{\mathrm{pred}} - \overline{\mathrm{GHI}^{\mathrm{pred}}})}{\sqrt{\sum_{i=1}^{N}(\mathrm{GHI}_i^{\mathrm{act}} - \overline{\mathrm{GHI}^{\mathrm{act}}})}\sqrt{\sum_{i=1}^{N}(\mathrm{GHI}_i^{\mathrm{pred}} - \overline{\mathrm{GHI}^{\mathrm{pred}}})}} \tag{3}$$

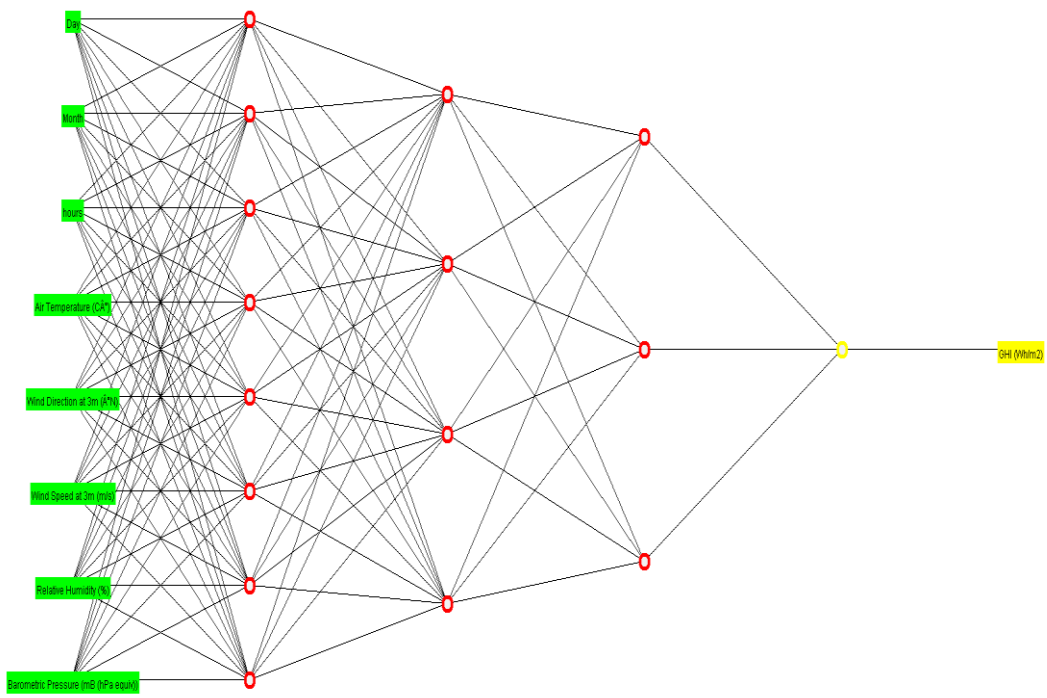

Figure 6.An ANN architecture for GHI predictions.

6. Result and Discussion

To test the three machine learning algorithms, we selected the city of Jeddah at the King Abdulaziz University Station (KAU, longitude 21.49604, latitude 39.24492). The database was divided into two stages: training and testing phases. The number of attributes, the percentage of training and testing, and the number of hidden layers and neurons were selected according to the accuracy determined by Root-Mean-Square-Error (RMSE), the mean absolute error (MAE), and the correlation coefficient (R). As explained in the previous section, there is no known solution for calculating the number of hidden layers and neurons; to overcome these challenges several tests were performed to select the optimum number of hidden layers and neurons. Once the number of layers and neurons has been set, additional runs were conducted to select the best percentage for training and testing. Our tests showed that 70%-30% for training and testing gives the best correlation. Table 1 shows the performance of GHI using Artificial Neural Networks (ANNs), Support Vector Machine (SVM), Random Forest, and Random Tree. A comparison between actual and predicted GHI reveals that Random Forest performed better than the rest of the machine learning techniques with a root mean square error (RMSE) and a correlation coefficient of (46, 0.99), respectively.
Figure 6 shows the comparison between the actual and predicted GHI using Random Forest. Figure 7 gives the linear regression between actual and predicted GHI values. Comparing these two figures with the GHI prediction and linear regression using SVM in Figure 8 and Figure 9 respectively reveals that Random Forest compares much better to experimental data.

Table 1. Performance of GHI using different Machine Learning algorithms

Machine Learning	**ANN**	**SVM**	**Random Tree**	**Random Forest**
R	0.98	0.78	0.97	**0.99**
RMSE	106	201	80	**46**
Relative Absolute Error	34%	57%	13%	**8%**

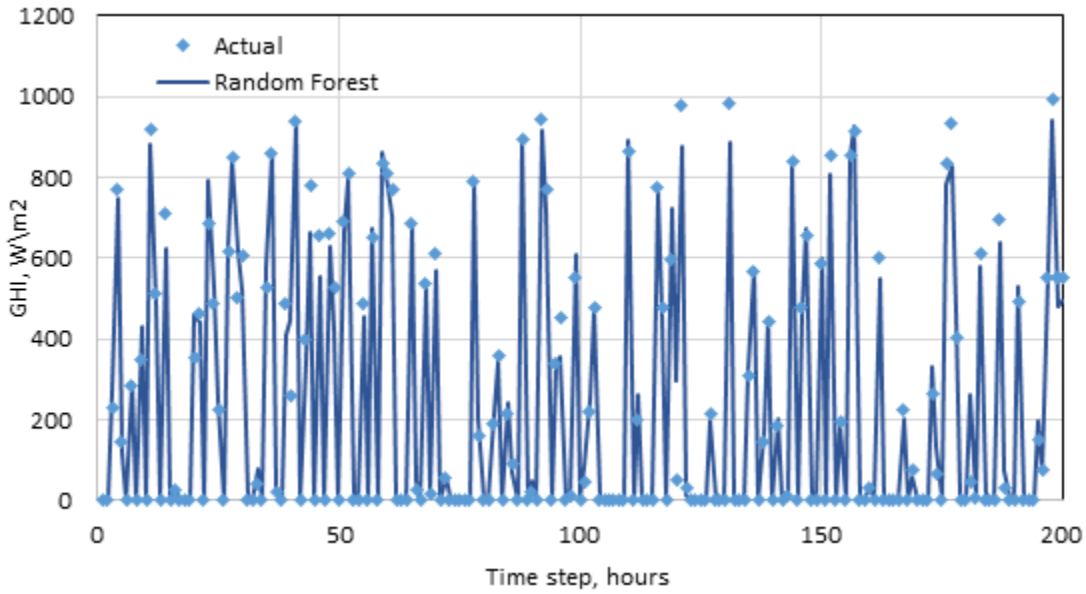

Figure 6. GHI prediction for the city of Jeddah at KAU Station, Jeddah, 2015, Random Forest algorithm

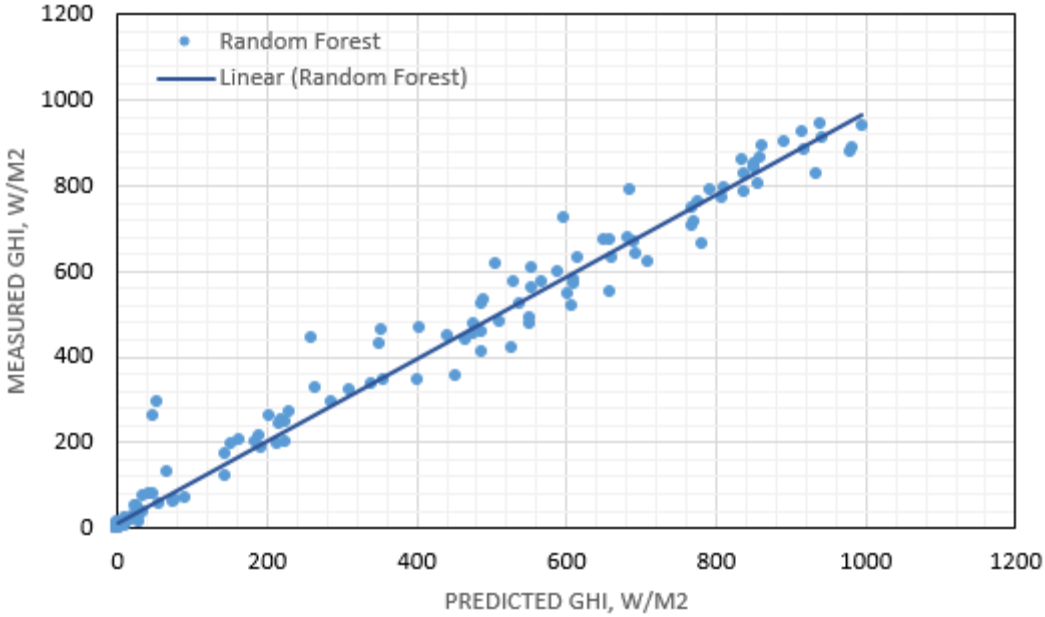

Figure 7. Actual and predicted GHI at KAU station, Jeddah, KSA, 2015, Random Forest algorithm

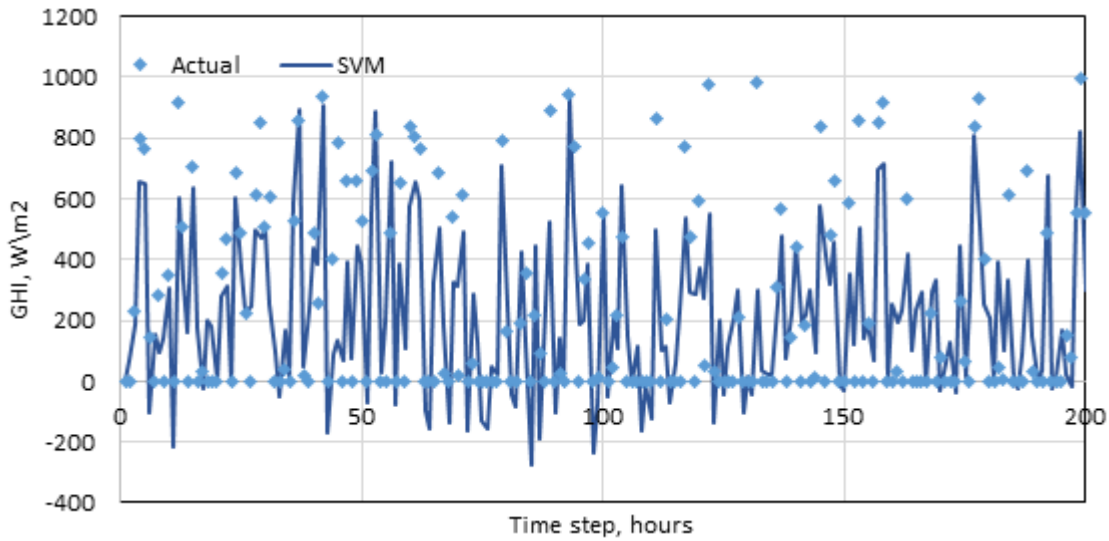

Figure 8. GHI prediction for the city of Jeddah at KAU Station, Jeddah, 2015, SVM algorithm

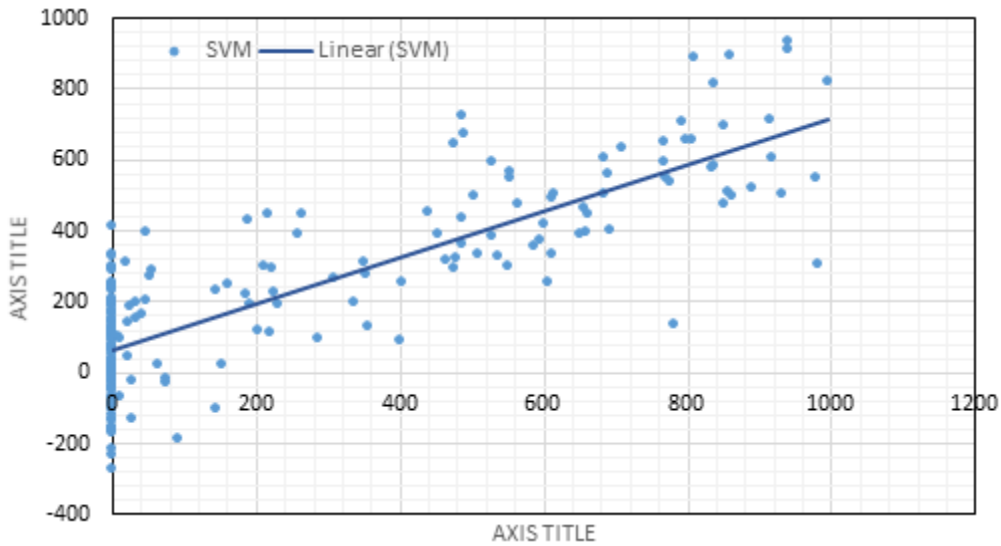

Figure 9. Actual and predicted GHI at KAU station, Jeddah, KSA, 2015, SVM algorithm

7. Conclusion

Due to its ever-increasing demand growth, the incorporation of solar energy into the grid becomes important. The present paper investigated four machine-learning algorithms, the ANN, SVM, Random Forest, and Random Tree. A comparison between actual and predicted GHI revealed that Random Forest performed better than the rest of the machine learning techniques with a correlation coefficient of 0.99. Data obtained from King Abdullah City for Atomic and Renewable was trained with different percentages, based on the RMSE and the correlation coefficient "R," the best result was obtained 70% - 30% for data training and testing gives. More work in underway to improve the RMSE to be more close to zero.

References

Alencar, D. B. D., Affonso, C. D. M., Oliveira, R. L. D., Rodríguez, J. M., Leite, J., & Filho, J. R. (2017). Different Models for Forecasting Wind Power Generation: Case Study. Energies, 10(12), 1976. doi: 10.3390/en10121976

Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

Ayyadevara, V. K. (2018). Random Forest. *Pro Machine Learning Algorithms*, 105–116. doi: 10.1007/978-1-4842-3564-5_5

Fratello, M., & Tagliaferri, R. (2019). Decision Trees and Random Forests. *Encyclopedia of Bioinformatics and Computational Biology*, 374–383. doi: 10.1016/b978-0-12-809633-8.20337-3

Greenpeace, EPIA, "Solar PV electricity empowering the world," Solar Generation 6, European Photovoltaic Industry Association EPIA, 2011.

IRENA, "Renewable energy market analysis," Retrieved May 2018 from http://irena.org/publications/2016/Jan/Renewable-Energy-Market-Analysis-The-GCC-Region.

K.A.CARE, 2017, Renewable Resource Atlas, King Abdullah City for Atomic and Renewable Energy (K.A.CARE), Saudi Arabia, https://rratlas.kacare.gov.sa.

KAPSARC Data Portal, (2017). Retrieved from https://datasource.kapsarc.org/explore/dataset/electricity-consumption-per-capita/table/?flg=ar&sort=year

KSA, Kingdom of Saudi Arabia (2017a). National Transformation Program. Retrieved from http://vision2030.gov.sa/sites/default/files/NTP_En.pdf

KSA, Kingdom of Saudi Arabia. (2017b) "Vision 2030. Retrieved May 2018 from http://vision2030.gov.sa/enKSA-NTP.

Lashin, A., Al Arifi, N., Chandrasekharam, D., Al Bassam, A., Rehman, S., Pipan, M., (2015). Geothermal energy resources of Saudi Arabia: Country update. World Geothermal Congress 2015. p. 15.

MEIM, Ministry of Energy, Industry and Mineral Resources (MEIM), (2019). Embarking on round two of Saudi Arabia's National Renewable Energy Program. Retrieved August 2019, from https://www.powersaudiarabia.com.sa/web/attach/news/Press-Release-080719-REPDO-Round2RFQ-EN.pdf

Mosavi, A., Salimi, M., Ardabili, S. F., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. (2019). State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. *Energies*, *12*(7), 1301. doi: 10.3390/en12071301

NREP, 2019, Saudi Arabia Renewable Energy Targets. Retrieved from https://www.powersaudiarabia.com.sa/web/attach/media/Saudi-Arabia-Renewable-Energy-Targets-and-Long-Term-Visibility.pdf.

Sanyour, O., Mofawaz, R., Basalem, G., Bensenouci, A., Brahimi, T., (2019). Design of a Solar Powered Air Conditioner". 10th, IEEE GCC Conference, April 21-23, 2019.

Shehata, M., Brahimi, T., Bensenouci, A., (2019) Design of a Solar-Powered University e-Library at Effat University Campus. IEEE ICIAES 2019, DUBAI, Industrial Applications of Engineering Sciences, Accepted for publication 2019, February 2019.

UN, (2015). United Nations, "Sustainable development goals. Retrieved August 2018 from https://sustainabledevelopment.un.org/sdgs

VAISALA, "Free wind and solar resource maps," Retrieved August 2018 from https://www.vaisala.com/en/lp/free-wind- and-solar-resource-maps.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., … Philip, S. Y. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37.

Zell, E., et al., "Assessment of solar radiation resources in Saudi Arabia," Solar Energy, Sep. 2015, vol. 119, pp. 422-438.