Predicting Renewable Energy Resources using Machine Learning for Wireless Sensor Networks

Satyam Bhatti, Ahsan Raza Khan, Sajjad Hussain and Rami Ghannam
James Watt School of Engineering, University of Glasgow, United Kingdom
Email: {s.bhatti.2, a.khan.9}@research.gla.ac.uk
{Sajjad.Hussain, Rami.Ghannam}@glasgow.ac.uk

***Abstract*—Wireless Sensor Network (WSN) nodes rely on batteries that are hazardous and need constant replacement. Therefore, we propose WSNs with solar energy harvesters that scavenge energy from the Sun. The key issue with these harvesters is that solar energy is intermittent. Consequently, we propose machine learning (ML) algorithms that enable WSN nodes to accurately predict the amount of solar irradiance, so that the node can intelligently manage its own energy. Our ML models were based on historical weather datasets from California (USA) and Delhi (India) for the period between 2010 to 2020. In addition, we performed data pre-processing, followed by feature engineering, identification of outliers and grid search to determine the most optimized ML model. In comparison with the linear regression model, the support vector regression (SVR) model showed accurate forecasting of solar irradiance. Moreover, it was also found that the models with time duration of 1 year and 1 month has much better forecasting results than 10 years and 1 week, with both root square mean error (RMSE) and mean absolute error (MAE) less than 7% for Sacramento, California, USA.**

***Index Terms*—Wireless Sensor Network (WSN), Internet of Things (IoT), Machine Learning (ML), Support Vector Regression (SVR).**

I. Introduction

Over the decade, there has been rapid growth in the field of the Internet of Things (IoT) due to its advanced connectivity with electronic devices. The IoT has transformed our lives to a large extent by efficiently connecting people across the world and turning our planet into a much smarter and more advanced globe [1]. In addition, IoT has the potential to profoundly enhance wireless networking technologies.

One of the subsets of IoT is the Wireless Sensor Network (WSN) which uses the combination of sensors to wirelessly interact and communicate with other sensor nodes. Moreover, the architecture of a WSN comprises a gateway node (central) connected with multiple sensor nodes (branches) to share information in the form of data packets from the transmitter to the receiving end [2]. The WSN with the help of sensor nodes, senses, gathers, processes, and transmits information such as weather sensing, health care industry, smart grids, robotics, and artificial intelligence.

The energy consumption in a WSN is due to sensing energy (sensors), computing energy (data processing) and communication energy (short radio-frequency circuit that performs data transmission and reception) [3]. The WSNs rely on batteries to feed power to the sensor nodes and gateway. However, battery capacity is limited due to a mismatch between the supply and demand. In addition, WSN nodes are installed in remote locations such as deserts, forests, war zones and in seas, [4] where human access is often restricted or limited. Moreover, frequent replacement of the battery is not possible in these remote locations [5] and thus, the energy management of the WSN nodes play a vital role in maintaining the prolonged operation with minimal investments. Therefore, we propose to develop energy harvesters that scavenge renewable energy from the surroundings [6].

Another limitation of energy harvesting is the intermittent nature of renewable resources. Though the sensor nodes are expected to consume less amount of energy, however, the main concern is the energy fluctuation and variable DC output from the solar cells. Not only does the energy is varying in sensor nodes but also it varies within the renewable energy resources due to their intermittent nature [7]. For instance, solar energy is available during the day only and wind energy varies with the wind speed. Hence, in our paper, we propose a model aiming to harvest renewable energy adequately from the environment overcoming the variations in renewable energy availability. Thus, we predict the amount of renewable energy using machine learning (ML) algorithms.

ML is a field of research that allows the machine to learn using past experiences and thus, train the machine to predict the future or possible outcomes. The ML comprises three types, supervised learning, unsupervised learning, and reinforcement learning. In our paper, we used supervised ML to predict the amount of solar irradiance generated on the WSN using historical data. We train and test our model using multiple solar irradiance parameters such as global horizontal irradiance, direct normal irradiance, ambient temperature, humidity, and latent heat of flux of the chosen location for ten years.

We performed feature extraction using correlation analysis of several parameters of solar irradiance such as global horizontal irradiance, ambient temperature, humidity, latent heat of flux, and normal direct irradiance. Further, we analysed our model by feature scoring of the multiple feature representing their dependency with each other using the correlation analysis and heat map. In addition, to estimate the hyper-parameters in multiple output support vector regression (SVR), we performed the grid search to find the optimised values of each hyper-parameter, i.e., (C, Gamma, and Kernel). Moreover, our paper represents a reduced computational complexity for

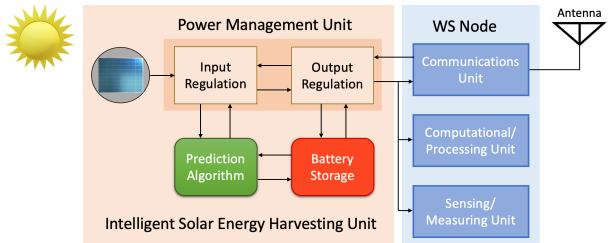

Fig. 1. The block diagram of the solar powered WSN and its constituent components.

determining the hyper-parameters for varied time series. In the last section, we made a comparison of the results from Linear Regression, Single Output SVR, and Multiple-Output SVR in terms of R2-Squared value, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) performances.

II. System Model

Figure 1 represents the block diagram of the WSN and its constituent components. The system scavenges energy from the surrounding, here from sun and the solar energy communicates with input regulation, battery storage, prediction algorithm and output regulation to smartly harvest solar energy. Further, the output from the power management unit is fed to the connected WSN node and antenna to share information with other sensor nodes. In addition, the next sections of the paper discusses the prediction algorithms for accurately forecasting the solar irradiance using ML techniques.

Figure 2 represents the working model of our proposed WSN architecture using ML techniques. The vector $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ is the input to the ML algorithm, which is used to train and test the model to predict the solar irradiance, $\mathbf{h} = \{h_1, h_2, ..., h_n\}$ is weight vector of the SVR hidden layers, and Y is the final forecasted solar irradiance that is the input to the main node of the WSN.

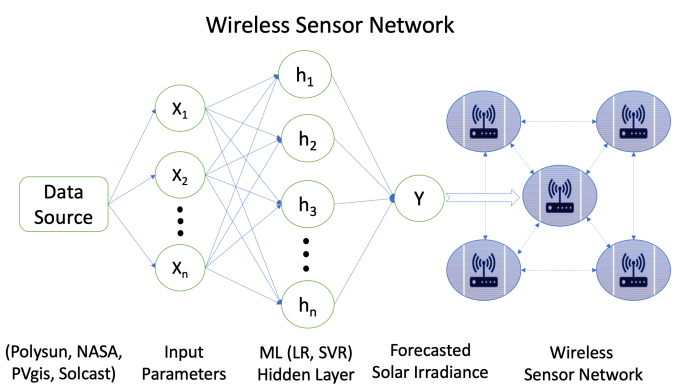

Fig. 2. Working Model of our proposed WSN architecture using ML techniques.

A. Data Processing

For our study, we chose the dataset for two locations, Sacramento, California, USA, and Delhi, India having global coordinates as (38.58, -121.35) and (28.58, 77.16,) respectively. The dataset for 2010 to 2020 is collected and divided so that the analysis is performed on the time interval of 1 week, 1 month, 1 year, and 10 years having an hourly time resolution. The dataset is accessed from the licensed version of the software Polysun [8] which provides solar irradiance values using the appropriate sensors installed at some particular locations in hourly resolution for a period of 1 year (randomly chosen year between 1996 to 2015).

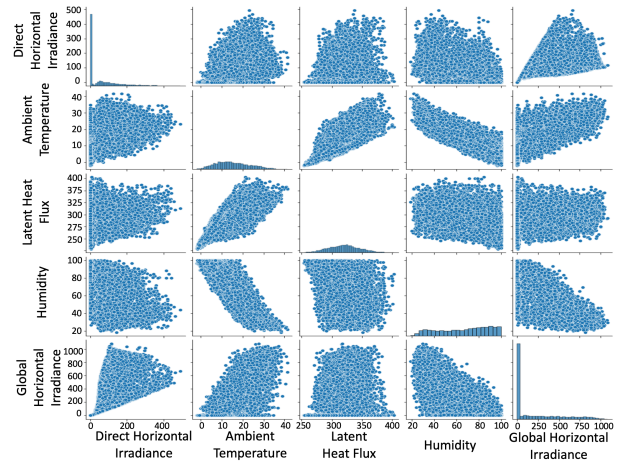

Fig. 3. Statistical representation for the solar irradiance dataset for California, USA.

Figure 3, depicts the historical graphs of the dataset and scattered data plot depicting the correlation of parameters to each other. The dataset consists of parameters such as global horizontal irradiance (Wh/m2), ambient temperature (ºC), latent heat of flux (W/m2), humidity (%), and direct normal irradiance (Wh/m2) for the location Sacramento, California, the USA for the year 2015.

B. Identification of Outliers in the dataset

For the SVR problems, the outliers play a significant role in determining the best fit line or hyperplane, thus it is important to identify the outliers in the dataset for each parameter. In our model, we considered four input parameters, global horizontal irradiance, latent heat of flux, ambient temperature, and humidity for the location California, the USA for 1 year. Figure 4 represents the box-plot of outliers in the SVR for each input parameter independently.

In addition, the parameter global horizontal irradiance has the maximum number of outlier which means that there is alot of noise in the dataset. Hence, it is useful to neglect the values of global horizontal irradiance which are greater than 300 W/m2. Moreover, there is some noise in the parameter latent heat of flux and ambient temperature, however, there are significantly lower values and thus, will not affect variance in the dataset.

C. Feature Engineering

Further, we perform feature engineering to determine the importance of each parameter with respect to each other. The

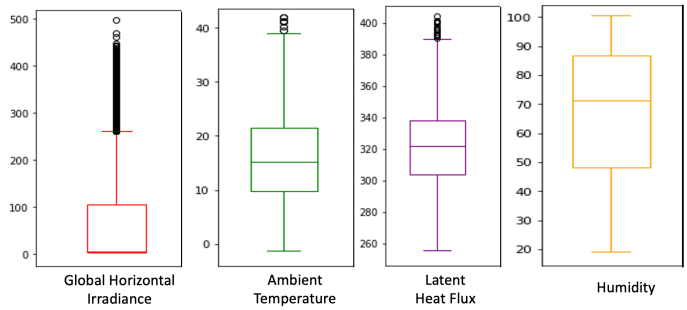

Fig. 4. Identification of outliers in the dataset for the parameters global horizontal irradiance, latent heat of flux, ambient temperature and humidity.

feature selection consists of correlation analysis and principle component analysis (PCA). For our study, we performed the correlation analysis using the heat map as shown in the figure 5. The direct normal irradiance has the maximum percentage of importance on output results, followed by ambient temperature and latent heat of flux. Further, the humidity has the least importance on the output results, hence, in our ML algorithm we will not consider humidity as an input variable.

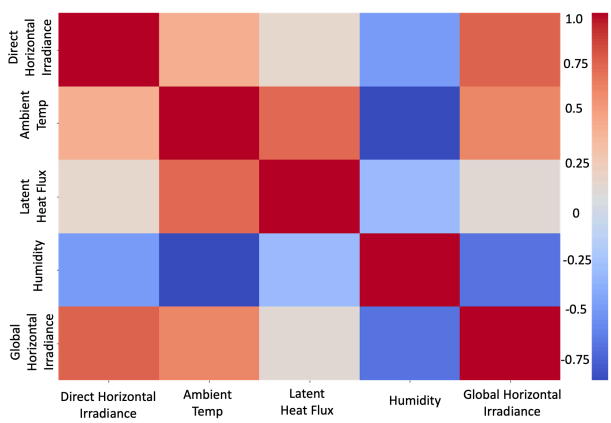

Fig. 5. Heat Map determining the correlation of the feature with respect to each other.

III. Model Training

A. Linear Regression Model

Mathematically, the linear regression model is a statistical approach of predicting the value of the cost function depending upon another variable. The linear regression model uses past experiences of the parameter as an input and predicts the outcome having a linear relationship between the cost function and output variables [9]. We implemented linear regression model because of the inter-dependency of multiple features on the prediction of solar irradiance.

For our model, we used linear regression with multiple features for forecasting solar irradiance. The 4 parameters were processed from the dataset of the location California, the USA for 1 year and further, divided into training (80%) and testing (20%) for the machine to learn in python.

B. Suppport Vector Regression Model

The Support Vector Machine (SVM) is an ML algorithm that consists of two types of paradigm, classification and regression problems. For our problem, we used the SVR algorithm to predict global solar irradiation. Like linear regression, the SVR is also classified as a Supervised ML algorithm that is used to determine the best possible fit line (linear) or a hyperplane (non-linear) containing the maximum points in a dataset [10].

Further, the optimized decision boundary for a line or a hyperplane can be calculated considering the most appropriate hyper-parameters values. In general, there are three hyper-parameters C, kernel, and gamma. First, the C, also known as the punishment factor of the error term, helps to evaluate the significance of outliers in the dataset. Second, the kernel is decided based on the dimensions of the dataset, for instance, if the decision boundary under consideration is linear or non-linear (radial basis function is used for multi-dimension kernel). Third, the gamma hyper-parameter for a non-linear function (uses radial basis function) influences the distance from a single train point in the dataset.

Similar to the linear regression model, we divided the dataset into the train (80%) and test (20%) and implemented the SVR model to forecast global solar irradiance. However, in SVR we also need to estimate the values of C, gamma, and kernel. Accordingly, we assumed the values of C to be in the range of 0.01, 0.1, 1, and 10. Also, the values of gamma were assumed to be 0.001, 0.01, 0.1, and 1 as well as for kernel, it was linear or rbf. The SVR uses the Grid Search function to determine the most optimized values for C, gamma, and kernel.

IV. Results and Discussion

A. Forecasting Solar Irradiance using LR Model

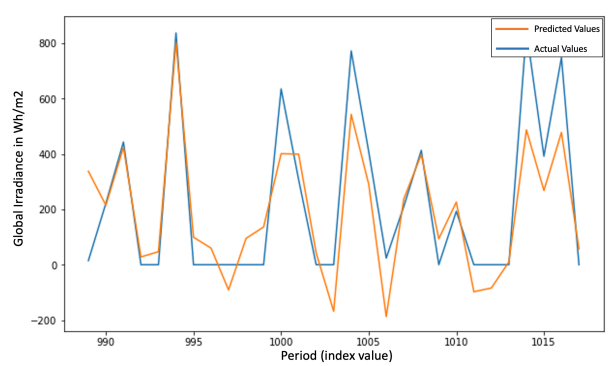

Fig. 6. Predicted solar irradiance using Linear regression for multiple parameters and zoom image of actual vs predicted curve.

Figure 6 depicts the plot of predicted global solar irradiance (orange curve) to the actual values (blue curve) of the solar irradiance for the location CA, USA. The overall accuracy of the model was 81.92%. In addition, we also implemented linear regression with multiple features for another location i.e. Delhi, India, and the overall accuracy was 87.30%.

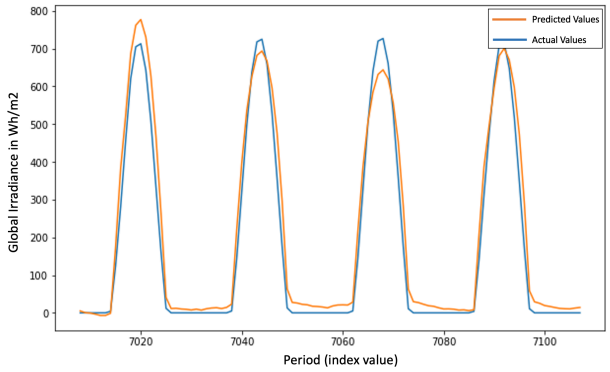

Fig. 7. Predicted solar irradiance using Multiple Parameter SVR and zoom image of actual vs predicted curve

B. Forecasting Solar Irradiance using SVR Model

Further, implementing the grid search for the calculations of optimized values of hyper-parameters, we obtained C = 10, Gamma = 0.001 and kernel = rbf. Further, incorporating these hyper-parameters as input to the SVR model, we divided the dataset into train and test of 80% and 20% respectively.

Likewise, the figure 7 represents the actual values vs forecasted solar irradiance using the SVR model. Here, we observed that the model has trained the dataset precisely, following the trend, and thus, forecasted the solar irradiance with an accuracy of 81.6% (also, known as the R-Square Performance value). The RMSE and MAE of the model is 6.9% and 3.1% respectively.

10 Years			1 Year		
Hyper-Parameter	**CA**	**Delhi**	**Hyper-Parameter**	**CA**	**Delhi**
C	10	10	C	10	10
Gamma	0.001	0.001	Gamma	0.01	0.01
Kernel	rbf	rbf	Kernel	rbf	linear
R-Squared (%)	39.68	40.21	R-Squared (%)	81.6	82.11
RMSE (%)	21.7	24.12	RMSE (%)	6.9	8.2
MAE (%)	13.07	17.9	MAE (%)	3.1	4.6
1 Month			**1 Week**		
Hyper-Parameter	**CA**	**Delhi**	**Hyper-Parameter**	**CA**	**Delhi**
C	10	10	C	1	10
Gamma	0.001	0.001	Gamma	0.01	0.001
Kernel	rbf	linear	Kernel	linear	linear
R-Squared (%)	76.47	84.13	R-Squared (%)	82.59	72.16
RMSE (%)	5.6	9.2	RMSE (%)	14.9	15.95
MAE (%)	2.4	6.8	MAE (%)	7.9	8.3

Fig. 8. Comparison of results based on RMSE, R-Squared value, and MAE.

Figure 8 consists of the optimized values of hyper-parameters (C, gamma, and kernel) calculated using the grid search. Also, the table includes the R-squared value, RMSE, and MAE for the locations Sacramento, California, USA, and Delhi, India for 10 years, 1 year, 1 month, and 1 week. According to the literature, the SVR model is expected to perform well if the R-square value is above 80% and RMSE, as well as the MAE value of the model, is less than 10% and 5% respectively. Therefore, the Multiple output SVR for 1 year and 1 month of California, USA, and Delhi, India has much better forecasting results as compared to the duration of 10 years and 1 week.

V. Conclusion

We conclude that an efficient method to power the WSN devices is by developing WSNs with energy harvesters that are capable to scavenge solar energy. Further, using the ML algorithms we accurately predicted the amount of solar irradiance, so that the node can intelligently manage its energy independently. We analysed data by executing data processing, identifying outliers and feature engineering to determine the most significant values of solar irradiances and respective parameters for two locations, i.e. California, USA and Delhi, India. Also, we performed grid search to find an optimized values of hyper-parameters, such as C=10, gamma = 0.001 and kernel = rbf for California, USA. Therefore, performing these optimization techniques helps to improve the overall prediction performance. Moreover, our ML model showed that the output results from the SVR model predicted the solar irradiance more accurately as compared to the linear regression model. In addition, the compelling results that we found is that the models with time duration of 1 year and 1 month has much better forecasting results than 10 years and 1 week, with both RMSE and MAE less than 7% for Sarcramento, California, USA.

References

[1] A. R. Sfar, Z. Chtourou, and Y. Challal, "A systemic and cognitive vision for IoT security: A case study of military live simulation and security challenges," in *2017 International Conference on Smart, Monitored and Controlled Cities (SM2C)*, 2017, pp. 101–105.

[2] J. Cabra, D. Castro, J. Colorado, D. Mendez, and L. Trujillo, "An iot approach for wireless sensor networks applied to e-health environmental monitoring," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2017, pp. 578–583.

[3] N. Alaoui, J.-P. Cances, and V. Meghdadi, "Energy consumption in wireless sensor networks for network coding structure and ARQ protocol," in *2015 International Conference on Electrical and Information Technologies (ICEIT)*, 2015, pp. 317–321.

[4] F. Tonolini and F. Adib, "Networking across boundaries: Enabling wireless communication through the water-air interface," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 117–131. [Online]. Available: https://doi.org/10.1145/3230543.3230580

[5] N. Khalil, M. R. Abid, D. Benhaddou, and M. Gerndt, "Wireless sensors networks for internet of things," in *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2014, pp. 1–6.

[6] K. S. Adu-Manu, N. Adam, C. Tapparello, H. Ayatollahi, and W. Heinzelman, "Energy-harvesting wireless sensor networks (eh-wsns): A review," *ACM Trans. Sen. Netw.*, vol. 14, no. 2, apr 2018. [Online]. Available: https://doi.org/10.1145/3183338

[7] S. Lata, S. Mehfuz, and S. Urooj, "Secure and reliable wsn for internet of things: Challenges and enabling technologies," *IEEE Access*, vol. 9, pp. 161 103–161 128, 2021.

[8] H. Elsaraf, M. Jamil, and B. Pandey, "Techno-economic design of a combined heat and power microgrid for a remote community in newfoundland canada," *IEEE Access*, vol. 9, pp. 91 548–91 563, 2021.

[9] H.-I. Lim, "A linear regression approach to modeling software characteristics for classifying similar software," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1, 2019, pp. 942–943.

[10] D. M. C. Nieto, E. A. P. Quiroz, and M. A. C. Lengua, "A systematic literature review on support vector machines applied to regression," in *2021 IEEE Sciences and Humanities International Research Conference (SHIRCON)*, 2021, pp. 1–4.