Solar Energy Prognostics using Deep Learning for Demand Side Management towards Affordable and Clean Energy

Himanshu Priyadarshi
Department of Electrical Engineering, Manipal University Jaipur

Abstract— Accelerated by gubernatorial incentivization, the abundance of solar energy harvestable in India presents a great opportunity for integrating solar photovoltaic (PV) systems with demand side management solutions towards the attainment of sustainable development goals (SDG) of affordable and clean energy. Sustainability in terms of profitability of such systems is dependent not only on their deployment, but also how well these solutions perform when passed to the test of consistency in terms of their reliability and performance. In order to act in tandem with energy storage management solutions with reliability and profitability in performance, importance of accuracy in predicting the solar energy available can hardly be overstated. In the experiment design associated with this manuscript, an openly available meteorological dataset from Jaipur has been used to assess the performance of deep learning-based regression techniques for solar insolation prognostics within the classical paradox of time complexity and speed of prediction. These techniques have been evaluated using key performance fidelity metrics including root mean square error (RMSE), mean absolute error (MAE), and speed of training. Statistical distribution of residuals and prognostic fidelity have been employed for graphical depiction of the techniques. The results highlight the trade-off between accuracy of prognostics and time complexity for choosing the technique based on application-specific requirements.

Keywords— Solar energy, Machine learning, affordable and clean energy, sustainable development goals(SDG)

1. Introduction

The call for action in the global context for transitioning to affordable and clean energy(SDG#7) sources has become increasingly imperative in pursuance of sustainable communities (SDG#11) with energy sovereignty. Environmental pollution and consequent public health issues as well as geopolitical instability interdigitated with the utter dependence on fossil fuels has acted as a wake-up call to the blind-folded patterns of hedonistic lifestyle. In response to this the dire need for climate action(SDG #13) has been realized for improving the quality of life on land(SDG#15) which is invariably interspersed with quality of life on water(SDG#14). Although the United Nations came into existence in 1945, and was supposed to be the watchdog for peace all over the world after Second World War, the world has continued to be pieced-up due to myopic technological practices with environmental pollution related casualties remaining very high. Realizing the need for tackling the menace created hedonistic and environmentally-malign technologies, the United Nations has realized that peace, justice and strong institutions(SDG#16) cannot be achieved when the technology does not respect the sanctity of nature, and hence has been trying to focus on environmentally-benign technology, which requires partnership at local as well as global level(SDG#17). Energy is the capacity to do work, and hence the energy technology is the prime propellant for economic development. Most unfortunately, the human civilization has been abjectly dependent directly as well as indirectly on most obnoxious fossil fuels for catering to energy needs. When the source for economic development itself is so malign, any industry, innovation or infrastructure(SDG#9) will be baffled in its hopes for attainment of decent work and economic growth(SDG#8). Apart from that potable clean water and sanitation(SDG#6) will be totally elusive if the environment is not clean. Aspirations of reduced inequalities(SDG#10) and attaining gender equality(SDG#5) is not possible in such a state of delirium caused deluded attempts for fossil fuel exploration [1]. When there is no consideration for environmental protection in spite of proven track record of disasters due to fossil fuels, setting goals of good-health and well-being(SDG#3) seems to be in sharp contradiction between the walk and the talk. Many crimes in the world can be attributed due to poverty and hunger, and hence the goals of no poverty(SDG#1) and zero hunger(SDG#2) are certainly noble. In order to attain these noble goals, quality education(SDG#4) must be rooted in empathy for planet and its environment.

In the previous centuries, the Indian subcontinent has been hailed as the golden bird for its unrivalled economic sovereignty with all the endeavour for economic development with empathy for planet and the environment. This economic sovereignty was coming from the abundance and diversity of natural resources this sub-continent has been blessed with [2]. It is not that the abundance and diversity has been withdrawn by Providence, rather it is still there, however the economic degradation has been due to mismanagement of the available resources. The abundance of renewable energy resources is a case in point, and by properly harvesting the solar energy, storing it using energy storage technology, and utilizing it prudently will very easily emancipate the country from dependence on crude oil. The actual potential of solar energy can be unleashed only when the solar photovoltaic panels are ably assisted by capable energy storage systems in-built with intelligence.

The current technology trends in energy storage has again been deluded due to abject dependence on lithium-ion batteries which have emerged as the dominant choice for mobility applications, however their use in stationary, grid-scale or residential energy storage poses significant challenges. Issues like high capital expenses, safety risks, and material availability restrict their scalability for large-scale applications. In all applications, the capable energy storage systems must be characterized by affordability, longevity, and safety. Hence, the development and optimization of indigenous energy storage strategies are essential for ensuring the economic viability of renewable energy infrastructure. Transitioning from internal combustion engine vehicles to electric vehicles in terms of mobility or from crude oil to lithium in terms of fuelling requirements is not going to solve the problem.

Energy security is foundational to societal development, as the availability of reliable, affordable power is directly proportional to a society's economic strength and the well-being of its citizens. Substantial energy reserves are mandatory for achieving the United Nations' Sustainable Development Goals. Either it is crude oil or lithium, the dependence on fossil fuels across many economies has led to dire consequences, including air and water pollution, climate change anomalies, and socio-political conflicts for resource dominion. With growing awareness of these dangers, the global energy dialogue has shifted toward decarbonization and decentralization, with renewable energy penetration and distributed generation (DG) playing pivotal roles. For an energy storage system to be hailed as capable, it must be indigenous apart from being durable and safe to handle.

However, the stochastic nature of renewable energy sources like solar and wind, present critical operational challenges. Solar energy harvested, for instance, is influenced by a range of determinable factors such as time of day, geographical location, and panel orientation, as well as unpredictable like temperature, humidity, cloud cover, and wind speed. These stochastic factors necessitate the deployment of energy storage systems (ESS) that can buffer the mismatch between generation and consumption. Moreover, intelligent management of these storage systems becomes unavoidable for enabling demand-side response, optimizing grid infrastructure, and reducing the need for additional power system investments. Smart energy storage solutions do support multiple applications like load levelling, peak shaving, emergency backup, and financial arbitrage. By accumulating excess energy generated during low-demand periods and releasing it in the wake of high-demand intervals, storage systems contribute to grid stability and mitigate reliance on costly peaking power plants. In markets with time-of-use pricing models, energy can be stored when selling prices are low and discharged when selling prices are high, offering benefits to both utilities and consumers.

The fast-paced evolution of the smart grid has further facilitated these innovations by incorporating real-time data collection, advanced analytics, automation, and control. The success of what has been mentioned in foregoing hinges on the prognostics. The ability to accurately forecast insolation allows for proactive adjustment of load schedules and energy storage re-charging cycles. This, enhances the reliability, efficiency, and profitability of solar energy systems by integration with decentralized generation sources like rooftop solar installations. In the context of prognostics, machine learning techniques made robust due to high level of accuracy by deep neural networks in the architecture powered by computational capability have emerged as powerful tools for modelling non-deterministic relationships between meteorological variables and solar energy availability. Closed-form models are not capable to deal with stochastic nature of weather-related factor. However, deep neural network architectures are capable of deciphering intricate patterns from variegated datasets and hence these are perfectly suited for the forecast of solar radiation. The integrated capability of deep learning and solar energy technology has been projected to address the most pressing issues of our time, including energy dependency, environmental pollution, and economic disparity. By harnessing the synergy of predictive data analytics as well as solar energy, unprecedented levels of efficiency and resilience can be achieved in demand side management, clearing the roadblocks in the journey towards affordable and clean energy.

In this manuscript, an experiment has been designed to assess the performance of pertinent deep learning-based regression techniques [3-9] for predicting solar insolation, by using an open-source weather dataset from Jaipur, India. Due to its high solar irradiation exposure and rapidly growing energy demands, Jaipur serves as a representative location for this research. With the focus on the likes of techniques like Gaussian Process Regression(GPR), Decision Tree Regression(DTR) and Support Vector Regression(SVR) [10-16], performance metrics such as Root Mean Square Error (RMSE), Euclidean Error or Mean Absolute Error (MAE), and time complexity(quantified by training time) using MATLAB[15-21].

The importance of this experiment lies in its utility to the intelligent energy storage management devices' management systems. A responsible control strategy with accurate solar forecasting can enable seamless integration of solar energy harvested in the broader scheme of distributed generation grid penetration(DGGP) by renewable energy sources. Although, it can be argued that DGGP is need only where grid infrastructure is not yet maturely developed or strained, but this argument is dangerous as exploration of fossil fuels has already been proved to be lethal for human health and existence on the planet. From an economic perspective, renewable energy storage systems are built on asset-lite model, which bears great significance for energy sector start-ups aiming to venture into the renewable energy market. Asset-lite business models emphasize minimal infrastructure requirements while maximizing productivity through innovative solutions and optimized platform design. For sustainability and profit, such business ventures depends on experiential data-driven energy management systems flexible enough to adapt to changing conditions and provide actionable insights for all the stakeholders of energy market including producers as well as consumers. Thus, this work contributes to the ongoing efforts toward clean energy transition by giving a comparative analysis of pertinent predictive modelling techniques for solar energy harvetsing. It highlights the need for integrating deep learning algorithms into energy management systems for the sake of operational effectiveness. Hopefully, such research endeavours can accelerate the global shift toward affordable and clean energy—a SDG that apart from being technically feasible is also ethically imperative in the context of a rapidly warming planet.

2. EXPERIMENTAL DESIGN

Traditionally, Jaipur has been known for its tourism stemming from the following reasons: (i) the heritage Pink city is known as a replica city of Vrindavan, due to its intimate historical connection and patronage with internationally known holy city of pilgrimage, Vrindavan; (ii) the heritage fortified city also houses a replica of Kaashi, the holy city situated in Benares in the state of Uttar Pradesh; (iii) the forts in Amber, Jaigarh, Naahargarh, as well as the Jantar-Mantar (the astronomical observatory) have been the object of research. These forts exhibit models of astronomy, meteorology, as well as rainwater harvesting. However, apart from tourism in the heritage city as well as the fortification, the district of Jaipur has been consistently expanding in the past 12 years in terms of population. In order to cater to the needs of the subjects, the government has invested heavily in the smart city Jaipur, and solar energy harvesting is one of the thrust areas.

The primary source for data has been the openly available dataset online from Centre for Energy and Environment (CEE) at Malaviya National Institute of Technology (MNIT) Jaipur. This meteorological data can be acquired by using a combination of onsite measurements, collaborative partnerships, and openly available datasets. Using instruments such as pyranometers, pyrheliometers, and so on on real-time, location-specific data on insolation and related weather parameters like temperature, humidity, and wind speed can be curated. In order to build a bigger and variegated corpus collaboration with stakeholders such as industry partners, government bodies is highly desirable. The integrated development environment(IDE) for coding and regression analysis used in this experiment is MATLAB. Time-series data has been imported, preprocessed(including treating missing keys, normalization, and train-test subset partitioning). Regression models such as Linear Regression, Support Vector Regression (SVM), Decision Trees Regression, and Gaussian Process Regression have been for prognostics. K-fold cross-validation has been used for robustness and overfitting avoidance. Error metrics such as Euclidean or Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) to quantify prediction performance. Residual plots and error distributions have been used generated for graphical description of performance.

3. RESULTS and DISCUSSION

Computational burden is associated with handling big data, particularly when random-access memory (RAM) comes at a price. When the data to be loaded on RAM exceeds its space, or when the processor is taxed by competing concurrent processes, it leads to page fault and other issues. The page faults may cause process execution to become sluggish failing for execution within a reasonable time frame. To grapple with this., dimensionality reduction technique of Principal Component Analysis (PCA) has been employed. While training regression models on big datasets using deep learning algorithms, applying PCA helps to overcome RAM limitations and improve time complexity. Grounded in the Eigenvalue Theorem, this technique allows for the retention of critical data characteristics and simultaneously computational overhead. This makes the learning process scalable and robust even with limited hardware resources.

Regression using Decision trees: In this approach, the heavier dataset is recursively split into lighter subsets based on the feature that provides the optimal partition for a given loss function, typically the Euclidean Error Squared. At each node of the tree, the algorithm assesses all possible splits for all features and takes the one that minimizes the variance of the target variable in the subsequent subsets. This continues till the attainment of a stopping criterion such as maximum depth of tree(DOT), a minimum number of PLO(Per Leaf Observations), or the value of error metric. Each terminal node, or leaf is required to hold the average of the target values for the samples associated with that region, that is the prediction for any new input that reaches the same leaf. Decision trees for regression are highly interpretable and are capable of model convoluted relationships. On the flip side these are prone to overfitting, as the tree grows too deep. Pruning (posteriori or apriori), cross-validation, and ensemble methods like Random Forests and Gradient Boosted Trees may be used to avoid overfitting due to high variance.

The results regarding the faithfulness of the predictions with respect to the ground truth along with the distribution of residuals for the bagged trees as well as boosted trees variant has been shown in figure 1.

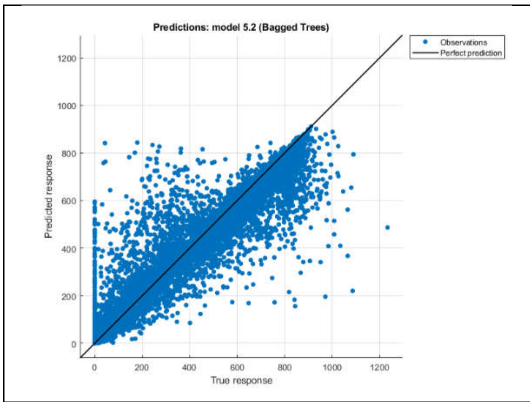
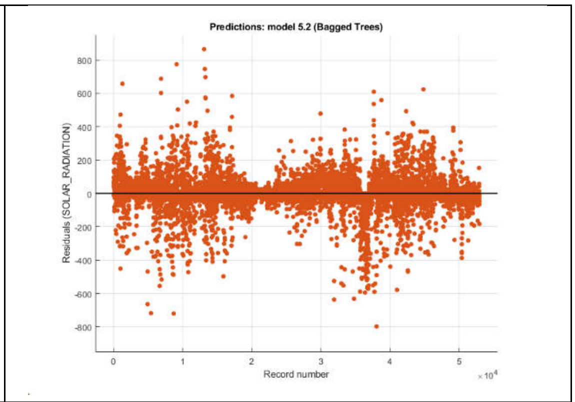
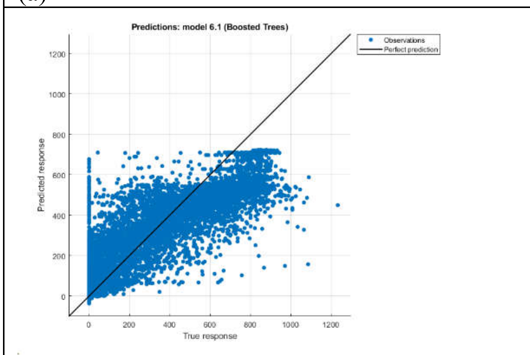
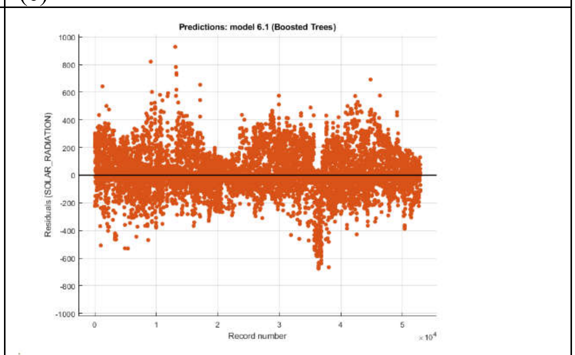
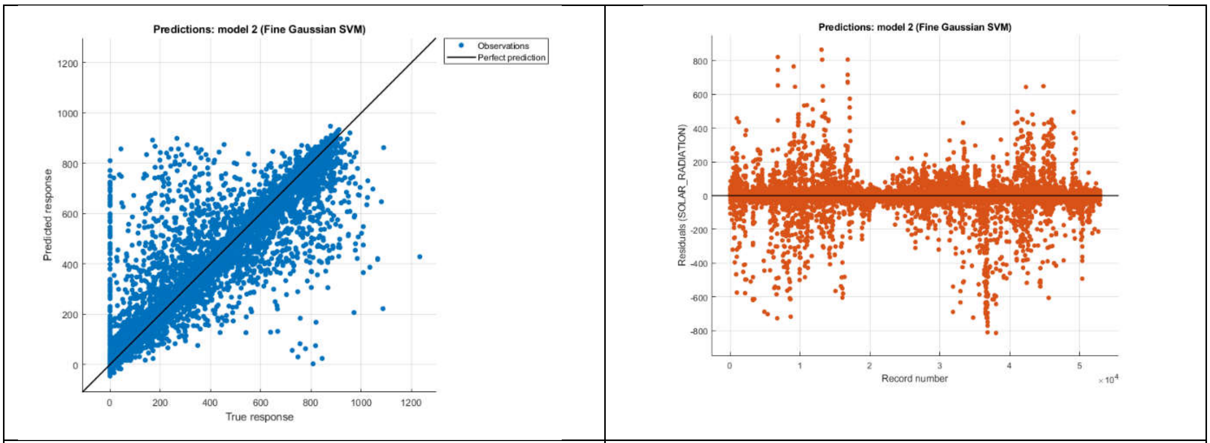
	
(a)	(c)
	
(b)	(d)

Figure 1: Graphical depiction of the solar energy received as predicted by Bagged Trees method (a) as well as Boosted Trees method (b) against recorded response, along with the distribution of residuals for Bagged Trees method (c) as well as Boosted Trees method (d) against record number.

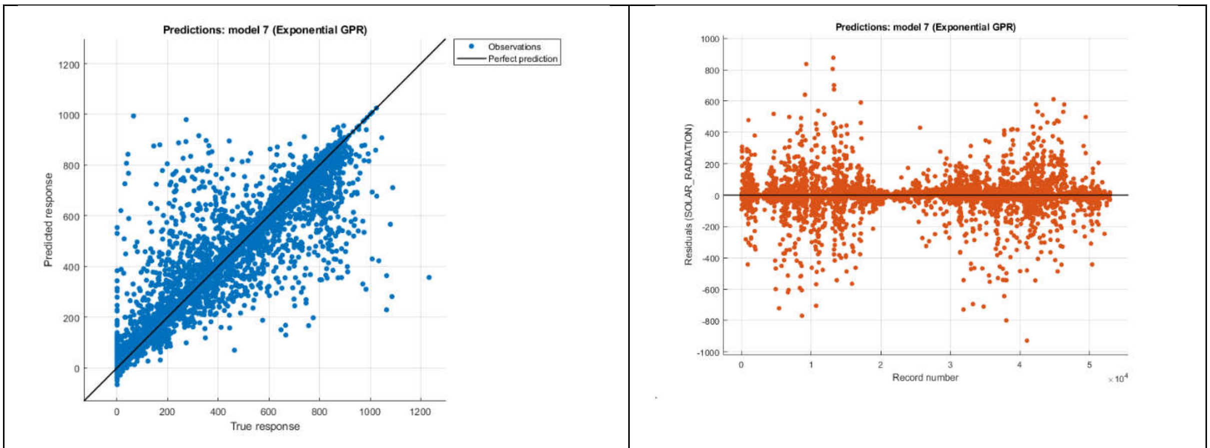
Regression using Fine Gaussian Support Vector Machine: In contrast to linear models[8-9] that assume a straight-line relationship, the Fine Gaussian SVM uses a radial basis function (RBF) kernel for transforming it into a higher-dimensional space making it easy to fit intricate patterns and to be more sensitive to subtleties of data. This can improve accuracy when the underlying function is highly nonlinear. This makes Fine Gaussian SVM especially suited for datasets where the response depends on complicated interactions between predictors or in scenarios where traditional regression models may underperform. The fidelity of the predictions in the context of this fine gaussian support vector machine regression as well as the associated distribution of residuals has been shown in figure 2 (a) and figure 2(b) respectively.



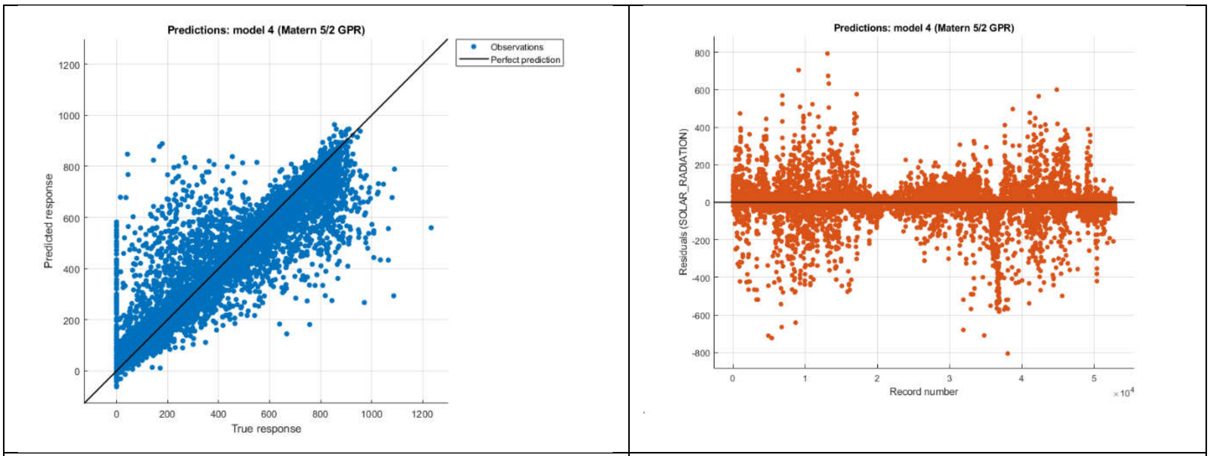
(a) (b)

Figure 2: Graphical depiction of the solar energy received as predicted by Fine Gaussian Support Vector Machine(SVM) method (a), along with the distribution of residuals (b) against record number.

Gaussian Process Regression: This method has been used to model non-linear complex relationships by treating the regression task as a distribution over possible functionsfitting a single model. Exponential kernel has been selected as it can capture local variations in the data. Although the fitting may be rougher , but more flexible. Being computationally costly, it is used only in situations where accurate modelling of uncertainty is important and where the dataset is comparatively smaller. Additionally, GPR has been favoured due to non-parametric nature avoiding rigid assumptions about the model structure, ability to quantify prediction confidence—advanatgeous in experimental design involving Bayesian optimization, and risk-sensitive decision-making. The fidelity of the predictions as well as the distribution of residuals has been shown in figure 3 (a) and figure 3(b) respectively. The use of Matern Gaussian Process Regression (GPR) has been motivated as it provides a flexible and interpretable way to model functions with variegated smoothness. In contrast to the commonly used squared exponential kernel, the Matern kernel includes a parameter for controlling the roughness of the function helping to capture real-world phenomena even with abrupt changes. This has made Matern GPR particularly valuable for geostatistics, environmental modeling, and other engineering applications, where data often exhibit spatio-temporal correlations with irregularities. Moreover, it balances complexity and generalization of the model, improving prognostic accuracy in cases where smoother kernels might fail to reveal the actual underlying structure of the data.



(a) (b)



(c) (d)

Figure 3: Graphical depiction of the solar energy received as predicted by Exponential Gaussian Process Regression method (a) as well as Matern 5/2 Gaussian Process Regression (b) against recorded response, along with the distribution of residuals for Exponential Gaussian Process Regression method (c) as well as Matern 5/2 Gaussian Process Regression (d) against record number.

The table shown below summarizes the results:

Table 1: Comparison of different deep learning algorithms (DLA) for insolation prediction on the basis of root mean square error (RMSE), mean absolute error (MAE), prediction speed, and training time.

Model type(AV =195, SD = 276) (PCA explaining 95% variance)	Tree	Fine Gaussian SVM	Gaussian Process Regression : Matern GPR	Bagged trees	Boosted trees	Exponential GPR
RMSE	103.33	97.348	89.37	86.869	133.98	71.33
R-squared	0.86	0.88	0.90	0.90	0.77	0.93
MAE	41.619	44.447	41.301	40.148	82.554	22.43
Prediction speed(obs/sec)	250000	3500	750	37000	210000	800
Training Time(sec)	16.306	145.74	1840	22.235	6.7656	3454.6

These results show the classic trade-off between time complexity and prediction accuracy in deep learning algorithms. Gaussian Process techniques has been apparently superior in terms of accuracy, as exhibited by graphical analysis of residual plots which have greater symmetry around the origin and fewer outliers. However, as far as training speed is concerned, bagged and boosted ensemble decision tree methods has been reported herein to be much faster due to their lower time complexity stemming from divide and conquer approach of the decision trees.

4. Conclusion

In this work on forecasting solar radiation, six different deep learning algorithms has been implemented for regression analysis. The effectiveness of each technique has been visually assessed through graphical analysis of their regression training. The findings bear an element of a trade-off between speed and accuracy among the choice of algorithms. Gaussian Process Regression (GPR) has been reported herein to have the highest accuracy in predicting solar radiation. This superior performance has an additional cost of a significant increase in computational time, making it less dependable for applications if speed is an overweighing factor. In contrast, the ensemble of decision trees has been proven to be much faster in training. Being computationally effective, its predictive accuracy was lower in comparison to GPR. Hence, the algorithm must be carefully considered based on the specific needs of the application in terms of required accuracy and the latency involved in prediction. There does not seem to be a universal model. Just like the choice of renewable energy to be harvested depends on the locally available resources, similarly the choice of machine learning strategy depends on the data. Although there are many cloud-based services for data analytics and data-sharing, the cost must be comprehensively considered. Moreover, inconsiderate expansion at the cost of environment must be avoided, and responsible consumers must be adequately incentivized for better demand side management.

5. References

[1] Priyadarshi, H., Ahmed, G., Mishra, D., Srivastava, A.K., McGee, R., Shrivastava, A. and Singh, K., 2024. Sustainable graphene-based energy storage device technology: Materials, methods, Monitoring and digital twin. *Critical Reviews in Solid State and Materials Sciences*, *49*(6), pp.1086-1119.

[2] Priyadarshi, H., Singh, K. and Shrivastava, A., 2021, August. Insolation prediction for reliable energy storage using deep learning algorithms. In *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)* (pp. 416-419). IEEE.

[3] Mehta, S., Misra, A., Singhal, A., Kumar, P. and Mittal, A., 2010. A high-performance parallel implementation of sum of absolute differences algorithm for motion estimation using CUDA. In *HiPC Conf* (Vol. 2, No. 4, p. 6).

[4] Pagalthivarthi, K.V., Mittal, A., Ravichandra, J.S. and Sanghi, S., 2007. Prediction of pressure drop in multi-size particulate pipe flow using correlation and neural network techniques. *Progress in Computational Fluid Dynamics, an International Journal*, *7*(7), pp.414-426.

[5] Gupta, R., Mittal, A., Singh, K., Bajpai, P. and Prakash, S., 2007, December. A time series approach for identification of exons and introns. In *10th International Conference on Information Technology (ICIT 2007)* (pp. 91-93). IEEE.

[6] Seshadri, G., Jain, R. and Mittal, A., 2010, February. Parallelization of Principal Component Analysis (using Eigen Value Decomposition) on scalable multi-core architecture. In *2010 IEEE 2Nd international advance computing conference (IACC)* (pp. 44-49). IEEE.

[7] Srivastava, A. and Mishra, N., 2024, July. Enhancing Photovoltaic Power Predictions with Limited Timeframe Data Using Neural Network. In *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)* (pp. 1-6). IEEE.

[8] Priyadarshi, H., Singh, K. and Shrivastava, A., 2022. Linear Regression for Insolation Prediction. In *Intelligent Computing Techniques for Smart Energy Systems: Proceedings of ICTSES 2021* (pp. 387-394). Singapore: Springer Nature Singapore.

[9] Priyadarshi, H., Singh, K. and Shrivastava, A., 2022. Green technology solutions for energy storage devices. In *Energy Conversion and Green Energy Storage* (pp. 117-132). CRC Press.

[10] Srivastava, A., Jha, S., Mishra, S., Parida, S.K. and Priyadarshi, H., 2024. Fault Prognosis in Smart Distribution System with Distributed Generation Using Fast Fourier Transform Assisted Machine Learning. *IETE Journal of Research*, *70*(9), pp.7415-7431.

[11] Srivastava, A. and Parida, S.K., 2022. Data driven approach for fault detection and Gaussian process regression based location prognosis in smart AC microgrid. *Electric Power Systems Research*, *208*, p.107889.

[12] Srivastava, A. and Parida, S.K., 2021. A robust fault detection and location prediction module using support vector machine and Gaussian process regression for AC microgrid. *IEEE Transactions on Industry Applications*, *58*(1), pp.930-939.

[13] Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J., & Seh, Z. W. (2020). *Predicting the state of charge and health of batteries using data-driven machine learning. Nature Machine Intelligence*

[14] Veronica Rabl et al,. Energy Storage Primer; IEEE Power & Energy Society, 2020.

[15] R. Patel and K. V. Pagalthivarthi, "MATLAB-Based Modelling of Power System Components in Transient Stability Analysis," in International Journal of Modelling and Simulation, vol. 25, no.1, pp. 43–50, 2005

[16] R. Patel and A. Mittal, Programming in MATLAB – A Problem Solving Approach, 1st ed., Pearson India ,2014

[17] A. Mittal and A. Kassim, Bayesian network technologies: applications and graphical models: applications and graphical models. IGI global, 2007

[18] Narang, V., Chowdhary, R., Mittal, A. and Sung, W.K., 2007. Bayesian Network Modeling of Transcription Factor Binding Sites: A Tutorial. *Bayesian network technologies: Applications and graphical models*, pp.300-318.

[19] A. Mittal and L.H. Cheong, "Addressing the problems of Bayesian network classification of video using high-dimensional features,", in *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 2, pp. 230- 244, March 2004

[20] Y. Li, M. Dong and R. Kothari, "Classifiability-based omnivariate decision trees," in *IEEE Transactions on Neural Networks*, vol. 16, no. 6, pp. 1547-1560, Nov. 2005

[21] S. Ertekin, L. Bottou and C. L. Giles, "Nonconvex Online Support Vector Machines," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 368-381, Feb. 2011.