Proceedings of the Canadian Society for Mechanical Engineering International Congress
31st Annual Conference of the Computational Fluid Dynamics Society of Canada
CSME/CFD2024
May 26–29, 2024, Toronto, Ontario, Canada

DEEP LEARNING PREDICTION OF SHORT-TERM HOURLY POWER OUTPUTS OF WIND AND SOLAR PV FARMS FOR REAL-TIME SMART BUILDING HVAC CONTROL

Jayasuriya Nandhagopal[1]*, Alan S. Fung[1]
[1]Department of Mechanical, Industrial & Mechatronics Engineering, Toronto Metropolitan University, Toronto, Canada
*jnandhagopal@torontomu.ca

***Abstract*—** Electrifying space heating of residential and commercial buildings through the use of Air Source Heat Pumps (ASHP) over Natural Gas/Fossil Fuel Furnace (NGF) has been enhanced by the development of Smart Dual Fuel Switching Systems (SDFSS) - a cloud-based IoT controller in buildings that enables automatic switching between ASHP and NGF based on energy cost optimization criteria. This aims to reduce the operation cost of ASHP while concurrently decreasing greenhouse gas (GHG) emissions. The operational cost and GHG emission of an SDFSS-enabled hybrid heating system could be further reduced by integrating the power output of wind farms with the SDFSS system. The power output data from 45 wind farms across Ontario were analyzed to determine energy generation patterns and relationships among input variables. The power output of three wind farms at Amarnath, Erieau, and Zurich in Ontario, Canada, was predicted in 1 to 6-hour time horizon using five machine learning algorithms: 1) Multiple Linear Regression, 2) MLPRegressor, 3) Random Forest Regressor, 4) Support Vector Regressor, and 5) Long Short-Term Memory. The time horizon and algorithm providing the best predictability were determined as a two-hour ahead prediction and the LSTM algorithm, rendering the model suitable for integration into the real-time SDFSS system.

Keywords-component; formatting; style; styling; insert (key words)

I. INTRODUCTION

De-carbonizing the society to achieve the Net-Zero emission by 2050 is the primary target of the Paris Agreement drafted by United Nations Framework Convention on Climate Change (UNFCCC). In resonance, the Canadian Federal Government's Pan Canadian Framework dictates the reduction of Canada's greenhouse gas (GHG) emission by 30% by 2030 and by 80% by 2050 [1, 2]. The deep examination of Canadian GHG emissions by sector shows that the residential sector is the third-largest contributor to national GHG emissions, accounting for a total contribution of 17%, of which 80% arises from space heating applications due to cold climatic conditions and large living space buildings, with an average house consuming 63% of its total energy for space heating by using natural gas as the predominant source of fuel for heating [1, 2, 3].

The switching of space heating from more polluting traditional natural gas fired furnace (NGF) to electricity driven air source heat pump (ASHP), reduces approximately 46-54% of annual GHG emission, while increasing the operational cost of space heating [4]. The high operational cost of ASHP could be reduced by using hybrid heating systems. The cloud based smart dual fuel switching system (SDFSS) reduces the operational cost and improves the efficiency of hybrid heating systems by enabling automatic switching between ASHP and NGF depending on multi-variable optimization process using the following variables time-of-use (TOU) pricing, fuel cost, weather forecast, equipment efficiencies and capacities [1, 3, 5]. There is a further opportunity to improve the efficiency of SDFSS system to reduce the operational cost of heating by integrating the heating system with renewable energy output from the local nearby solar PV farms and wind farms forming a local communal energy pooling. The more precise short-term prediction of energy output from local solar PV farms and wind farms is of paramount importance while integrating the renewable energy output to the SDFSS systems. The renewable energy should be predicted up to 6-hour head to be used with the SDFSS system.

The artificial neural network (ANN) models with an average correlation coefficient of 0.99 was developed using the weather data at a 2 min and 1 hour resolution to predict the outdoor temperature of residential building without requiring high precision sensors [6]. ANN model was developed using the Levenberg-Marquardt algorithms to predict the performance of heat pump systems in row-house with retrofits [7].

The wind speed sequence was decomposed into sub-sequences using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and the Least Squares Support Vector Machine (LSSVM) algorithm, optimized by the Artificial Bee Colony algorithm, was then applied to predict the wind speed [8]. Ensemble regressors and decision tree algorithms like k-nearest neighbors and Support Vector Regression (SVR) were developed to predict the energy output of wind farms, and it is found that the combination of decision tree with SVR outperforms the ensemble predictors [9]. Six different algorithms were developed to forecast the energy output of wind farms, namely XGBoost, Random Forest regressor, LightGBM, CatBoost, AdaBoost, and M5-Prime and among these algorithms, XGBoost produced the best performance in forecasting [10]. In predicting the wind farm's energy output for varying time horizons (1-hour, 1-week, and 12-month), six algorithms—Elastic Net Regression, Random Forest Regression, SARIMA, XGBoost, Prophet, and a combined Prophet and XGBoost model—were employed, with the SARIMAX model excelling in short- term forecasting and XGBoost delivering superior long- term predictability [11].

The present work focuses on the prediction of wind farms power output for varying time horizons (1-hour, 2-hour, 3-hours, 4-hours, 5-hours, 6-hours) using five algorithms: 1) MLPRegressor, 2) Support Vector Regression (SVR), 3) Random Forest Regressor, 4) Multiple Linear Regressor, and 5) Long Short-Term Memory (LSTM). The models were developed for three wind farms at Zurich, Erieau, Amaranth in the province of Ontario, Canada. Each model is developed using multiple test-train splits (70/30, 80/20, 60/35, 50/50) to determine the robustness of the model.

II. Machine Learning Architecture

A. Model stages

The machine learning and deep learning models developed in the paper to predict the hourly wind power output was developed in four stages. 1) Data processing and analysis, 2) Objective function definition, 3) Model training and performance evaluation, 4) Production AI.

B. Objective function selection

The objective function of the machine learning algorithms is predicting the wind farm energy output 1, 2, 3, 4, 5, 6 hours ahead.

C. Model training

The 6 objective functions of predicting the wind energy output are trained on each of the five-machine learning model chosen. The input dataset is randomly split into multiple test-train splits (70/30, 80/20, 65/35, 50/50) and trained using each of five algorithms for the given objective function as shown in Figure. 1

The mean correlation coefficient (r^2) for each algorithm across different test-train splits was computed. The algorithm with the highest average correlation coefficient (r^2) was selected as the best machine learning model for predicting the power output with a specific time delay, as shown in Figure 1.

Identify applicable sponsor/s here. *(sponsors)*

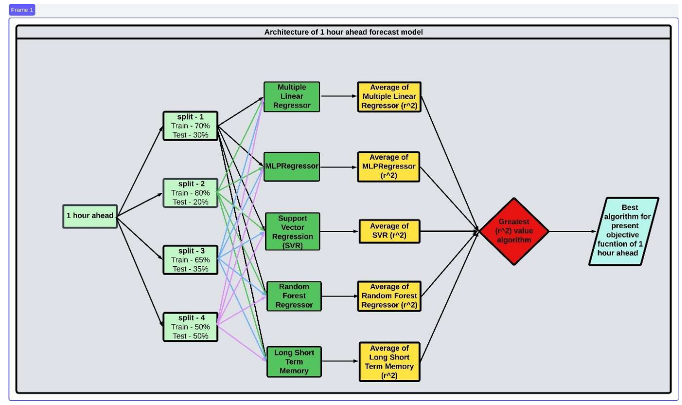

Figure 1. Architecture of single iteration of unique multilevel machine learning mode

III. Data Processing

A. Data collection and cleaning

In this stage, the historical hourly wind energy output data, along with the corresponding weather data for a period of three years and six months, were obtained through the Independent Electricity System Operator (IESO) portal and NASA POWER LARC, respectively, covering the period from July 2019 to January 2023. The data is then cleaned to remove inconsistencies, transformed into a standardized format, and stored in the database.

Out of 48 wind farms operational in Ontario, three wind farms shown in Table 1, are chosen for developing the machine learning and deep learning model.

TABLE I. POWER PLANT UNDER STUDY DETAILS

Generator name	Type of generator	Capacity	Latitude	Longitude	Operated by
Amarnath, ON	Wind	199.5 MW	44.10	-80.279	Transalta
Erieau, ON	Wind	99 MW	42.40	-82.183	Engie
Zurich, ON	Wind	100 MW	42.30	-82.161	Samsung renewable energy

B. Input metadata

The dimension of the feature space is denoted by m x n, where m is the number of records and n is the number of distinct features. The n for the dataset is 15 and it consists of the following parameters 1) Temperature, 2) Relative Humidity, 3) Precipitation, 4) Surface Pressure, 5) Wind speed, 6) Wind Direction, 7) All Sky Surface PAR Total, 8) Clear Sky Surface PAR Total, 9) All Sky Surface UVA Irradiance, 10) All Sky Surface UV Index, 11) All Sky Surface Irradiance, 12) Clear Sky Surface Irradiance, 13) All Sky Surface UVB Irradiance, 14) hour, 15) month.

C. Data manipulation – time series generator

The time series hourly power output prediction requires each hour's data to include the previous hours data to determine the influence of previous hour parameters on the present hour prediction. The particular hour of the day for which the prediction is made is denoted as P(t), and the input weather parameters for the prediction include Temperature(t), Windspeed(t), and more.

For a three-hour ahead power output prediction, weather forecast data from the previous three hours must be included, such as Temperature(t-1), Windspeed(t-1), Temperature(t-2), Windspeed(t-2), and so on increasing the number of features as shown in Figure 1. The data for 1, 2, 4, 5, 6 hours ahead predictions are transformed in the same manner as shown in Figure 1.

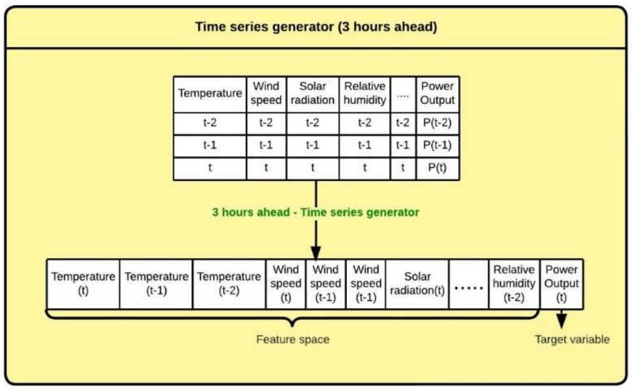

Figure 2. Time series generator for 3 hours ahead prediction

The dimension space of the input vector is m x n and the number of features for t - time ahead prediction is given by n.t

IV. Exploratory Data Analysis

The wind power output trend shows a stationary seasonality with high power output in winter months (October to April) and low power output in summer months (May to September) each year. The temperature vs wind power output at Amarnath wind farm shown in Figure 3, shows that with the increase in temperature the wind power output decreases.

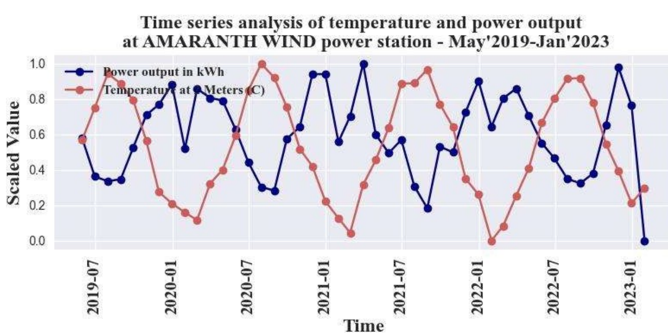

Figure 3. Timeseries analysis of wind farm power output and temperature between May 2019 - January 2023

The wind speed vs wind power output curves shown in Figure 4, shows that with the increase in wind speed the wind power output increases.

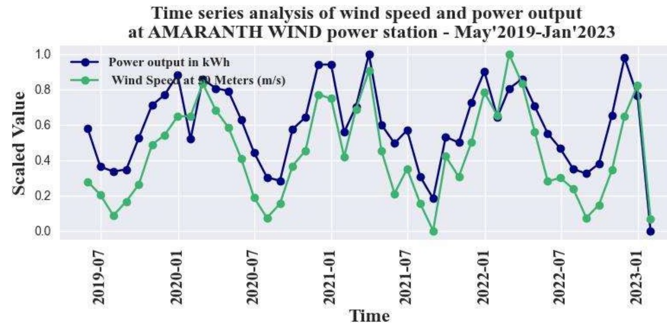

Figure 4. Timeseries analysis of wind farm power output and wind speed between May 2019 – Jan 2023

The solar radiation vs power output plot shown in Figure 5, shows that with the increase in solar radiation the wind power output decreases. This may be due the correlation between temperature and solar radiation which share a direct relationship among them and inverse relationship with the wind power output.

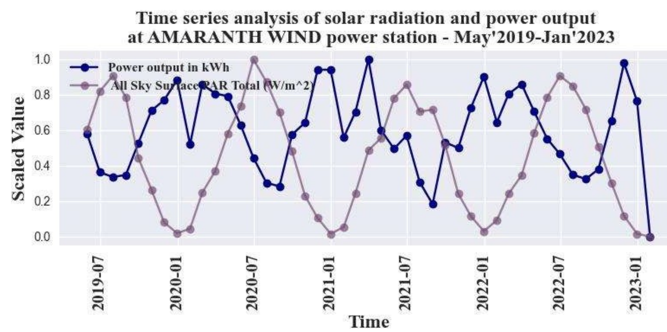

Figure 5. Timeseries analysis of wind farm power output and solar radiation between May 2019 – January 2023

The relative humidity vs power output plot shown in Figure 6, shows that with the increase in relative humidity the wind power output increase. This doesn't confirm any correlation between relative humidity and wind speed but shows both the variables exhibits a direct relationship with the wind power output.

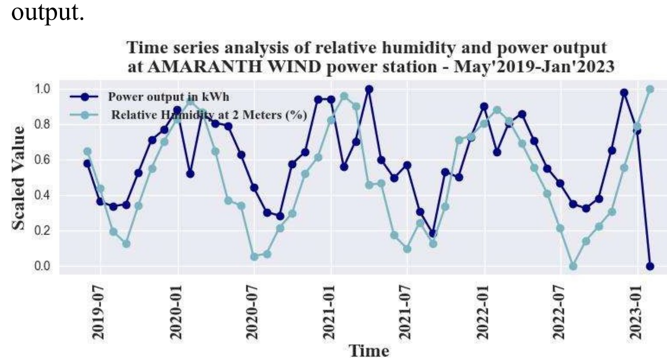

Figure 6. Timeseries analysis of wind farm power output and relative humidity between May 2019 – January 2023

The total monthly power output trend of the wind farm at Erieau is shown in Figure 7, along with the average monthly temperature, windspeed and total monthly solar radiation. The analysis indicates that during the winter season (October – April), wind farm power output is high, corresponding to high wind speed, relative humidity, demonstrating a direct relationship among these three variables. This pattern is consistent across all three wind farms. In addition, the analysis shows an inverse relationship between temperature and solar radiation with the power output from the wind farms.

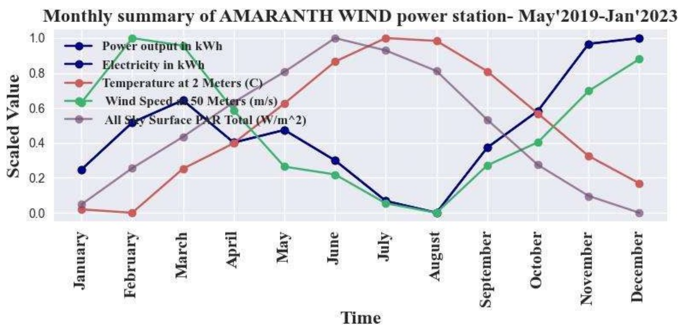

Figure 7. Monthly trend analysis of wind farm at Erieau

The hourly total power output of the Erieau wind farm exhibits a cyclic pattern as shown in Figure 8. During night hours when wind speed and relative humidity are high and temperature and solar radiation are low, the power output is at its peak. Conversely, wind power output is high when wind density and wind speed are elevated, while high temperature and solar radiation result in reduced wind density, leading to lower energy output during daylight hours when both are high.

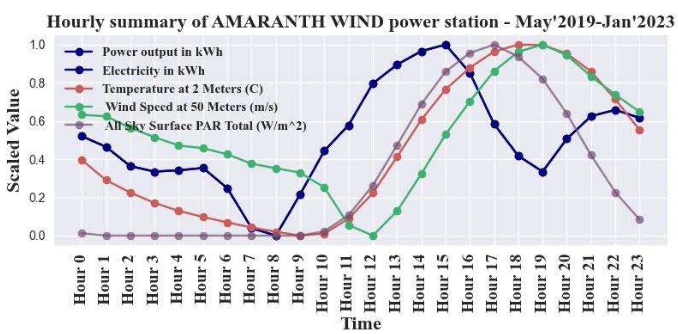

Figure 8. Hourly summary of wind farm at Erieau

The feature importance chart, as shown in Figure 9, illustrates the correlation between wind farm at Erieau's power output and its corresponding weather parameters. This chart highlights the parameters with the highest correlation to the target power output variable. Specifically, the wind speed and relative humidity exhibit the strongest correlations with the output power.

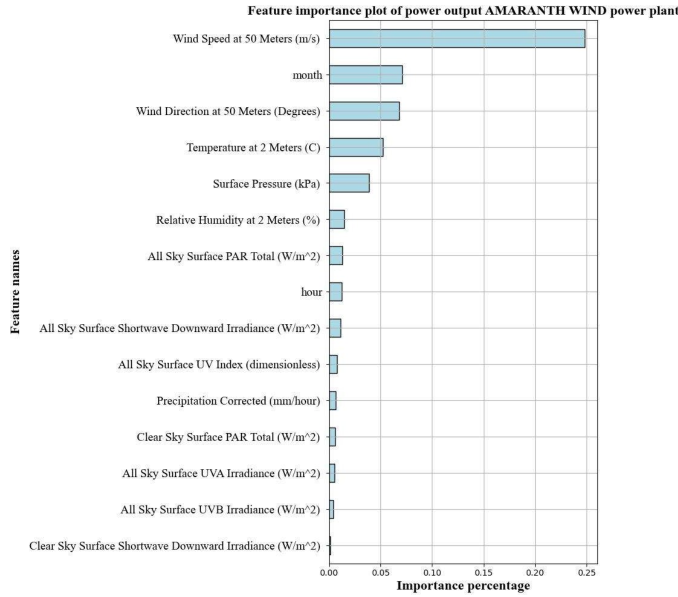

Figure 9. Feature importance of wind farm power output

V. MODEL PARAMETER SELECTION

A. *Multiple linear regression*

The hyperparameters used for multiple linear regression is learning rate: 0.1

B. *Random Forest regressor*

The hyperparameters used for Random Forest Regressor are n_estimatores: 100, n_estimators: 100, max_depth: 20, min_samples_split: 4, max_feature: 'auto', bootstrap: True, Criterion: MSE, max_samples: 0.7

C. *Multilayer Perceptron Regressor*

The hyperparameters used for Multi-Layer PerceptronRegressor (MLPRegressor) are hidden_layer_sizes: (60, 60), activation: relu, solver: Adam, learning_rate: constant, alpha: 0.001, max_iter: 1000

D. *Support vector regression*

The hyperparameters used for Support Vector Regression are kernal: rbf, C: 1, epsilon: 0.2.

VI. RESULT AND DISCUSSION

A. *Multiple linear regression*

The r^2 value of the multiple linear regression models were very less while predicting the power output for the current hour and the r^2 value increased while predicting the power output for the 2 – 6 hours ahead as shown in Figure 10. The higher r^2 value for hours ahead prediction is due to overfitting of the model and linear regression is not suitable for capturing the time lag information.

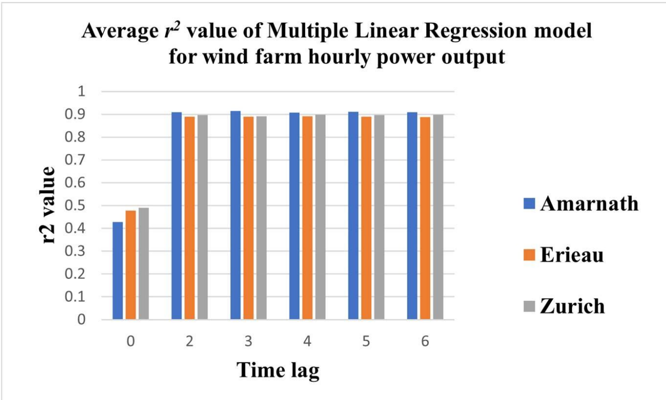

Figure 10. Multiple linear regression model's r^2 value for wind energy power output

B. MLPRegressor

The r^2 value of the MLPRegressor models were very less while predicting the power output for the current hour and the r^2 value increased while predicting the power output for the 2 – 6 hours ahead as shown in Figure 11. The average r^2 value for the model is 0.89 with current hour prediction models having an average r^2 of 0.65.

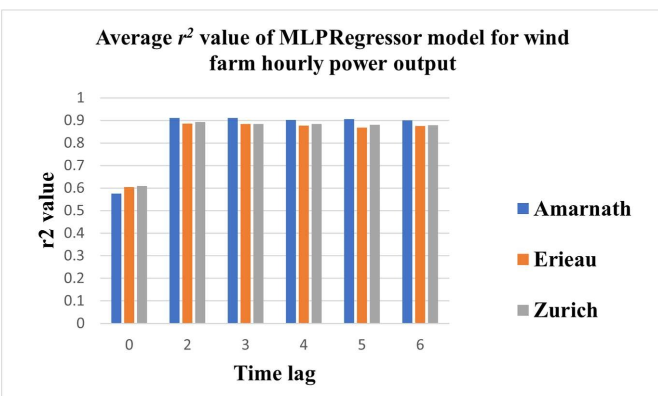

Figure 11. Average r^2 value of MLPRegressor models for predicting the power output in wind farms

C. Random Forest Regressor

The r^2 value of the random forest regressor models were high while predicting the power output for the current hour and the r^2 value increased while predicting the power output for the 2 – 6 hours ahead as shown in Figure 12. The average r^2 value for the model is 0.9 with current hour prediction models having an average r^2 of 0.7.

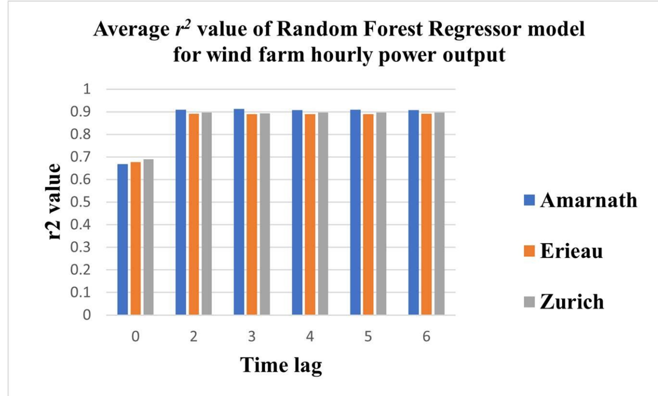

Figure 12. Average r^2 value of random forest regressor models for predicting the power output in wind farms

D. Support vector regression

The r^2 value of the support vector regressor models were less while predicting the power output for the current hour and the r^2 value of predicting the power output for the 2 – 6 hours ahead is also less when compared to other models as shown in Figure 13. The average r^2 value for the model is 0.8.

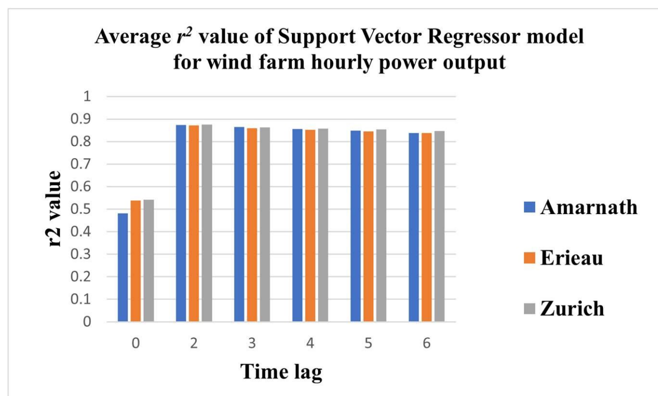

Figure 13. Average r^2 value of support vector regression models for predicting the power output in wind farms

E. Long short term memory

The r^2 value of the long short term memory models were highly accurate in predicting the power output for the current hour and the r^2 value of predicting the power output for the 2 – 6 hours ahead is also high when compared to other models as shown in Figure 14. The average r^2 value for the model is 0.9.

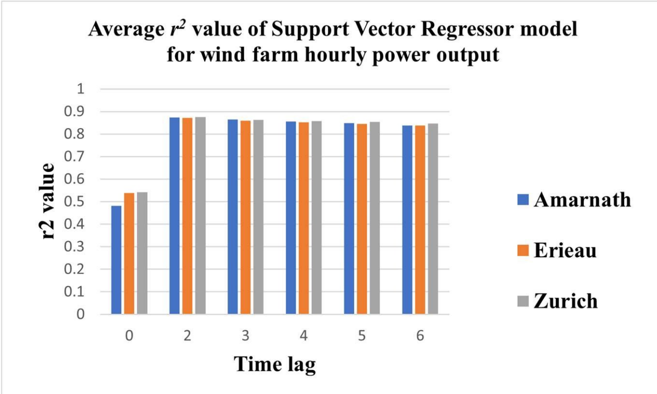

Figure 14. Average r^2 value of support vector regression models for predicting the power output in wind farms

VII. CONCLUSION

In conclusion, the hourly forecasting of wind farm energy output was more accurately and reliably predicted using the model trained using deep learning recurrent neural network long-short term memory model. The forecasting model using LSTM algorithm provided good accuracy in predicting power output between 2 and 6 hours ahead with an average r^2 value of 0.90 and average execution time of 6 minutes. The models trained using random forest regressor exhibited good r^2 value, but it was due to overfitting of the data. The SVR and MLPRegressor predicts the power output accurately if it's forecasting for current hour, when it's predicted between 2 to 6 hours the r^2 value gradually decreases. The execution time could be decreased by incorporation parallel processing algorithm which will reduce the load on single machine and distribute the processing load to multiple systems when required thereby resulting in near real time prediction with utmost? accuracy.

REFERENCES

[1] G. Demirezen and A. S. Fung, "Feasibility of Cloud Based Smart Dual Fuel Switching System (SDFSS) of Hybrid Residential Space Heating Systems for Simultaneous Reduction of Energy Cost and Greenhouse Gas Emission," Energy and Buildings, vol. 250, 2021, Art. no. 111237.

[2] G. Demirezen, K. Ullah, A. Rokn, and A. Fung, "Economical and environmental data analysis of hybrid HVAC system of air source heat pump and natural gas furnace for cold climate - Canada," in Proc. 13th IEA Heat Pump Conference 2020, Jeju, Korea, Apr. 26-29, 2021.

[3] D. Yu, K. Y. Tung, N. Ekrami, G. Demirezen, A. Fung, F. Mohammadi, and K. Raahemifar, "Proof of concept of a cloud-based smart dual-fuel switching system to control the operation of a hybrid residential HVAC system," in 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Macao, China, Dec. 1-4, 2019, pp. 1–5.

[4] A. M. Brockway and P. Delforge, "Emissions reduction potential from electric heat pumps in California homes," The Electricity Journal, vol. 31, no. 9, pp. 44–53, 2018.

[5] G. Demirezen, N. Ekrami, and A. Fung, "Monitoring and evaluation of nearly-zero energy house (nzeh) with hybrid hvac for cold climate canada," IOP Conference Series: Materials Science and Engineering, vol. 609, p. 062001, Sep 2019.

[6] G. Demirezen, A.S. Fung, and M. Deprez, "Development and optimization of artificial neural network algorithms for the prediction of building specific local temperature for hvac control," International Journal of Energy Research, vol. 44, no. 11, pp. 8513–8531, 2020.

[7] K. Ye, G. Demirezen, A. Fung, and E. Janssen, "The use of artificial neural networks (ann) in the prediction of energy consumption of air-source heat pump in retrofit residential housing," IOP Conference Series: Earth and Environmental Science, vol. 463, p. 012165, mar 2020.

[8] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in 2011 IEEE international conference on smart grid communications (SmartGridComm), pp. 528–533, IEEE, 2011 (17-20 October 2011, Brussels, Belgium).

[9] H. Wang, Y. Li, M. Xiong, and H. Chen, "A combined wind speed prediction model based on data processing, multi-objective optimization and machine learning," Energy Reports, vol. 9, pp. 413–421, 2023, presented at the 3rd International Conference on Power Engineering, Dec. 09–11, 2022, Sanya, Hainan Province, China.

[10] J. Heinermann and O. Kramer, "Machine learning ensembles for wind power prediction," Renewable Energy, vol. 89, pp. 671–679, 2016.

[11] C. Cakiroglu, S. Demir, M. Hakan Ozdemir, B. Latif Aylak, G. Sariisik, and L. Abualigah, "Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating shap analysis," Expert Systems with Applications, vol. 237, p. 121464, 2024.