Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan *Vol. 27 No. 4-2022*
ISSN:1813-1786 (Print) 2313-7770 (Online)

Machine Learning Substitute Modelling of Computational Fluid Dynamics Simulations for Wind Turbine Wakes

H. L. Ali[1], S. A. R. Gardezi[2], A. Bari[3], M. U. A. Khan[4], M. A. Intizar[5], M. Faisal[6]

[1,2,3,4,5,6]*Department of Mechanical Engineering, Bahauddin Zakariya University, Multan*

[2]asad.gardezi@bzu.edu.pk

Abstract- The growing need for sustainable energy sources has led to a focus on improving the power production capacity of wind turbines. The power optimization of wind turbines is chiefly hinged on the wind velocity which is mitigated by the wake and turbulence effects produced within turbine blades. Herein, computational fluid dynamics (CFD) techniques have been used for resolving wake effects nevertheless the computational method is expensive and tedious. In this context, a machine learning strategy named surrogate modeling was used to predict the reduced velocities inside the wake. These models were trained from a small data set attained from CFD simulations like input air velocities of 6 m/s , 9 m/s and 12 m/s. The machine learning surrogate models provided a data set that aided to find the wind velocity at any arbitrary point and helped in calculating required parameters in real-time without running costly CFD simulations with least error like mean absolute error of 0.000096 by GBR model. Additionally, contribute to power improvement and reliability of wind turbines and wind farm layout optimization. The findings are applicable for optimized performance of the wind turbine wakes at bench and commercial scales

Keywords-- Machine Learning; Computational fluid dynamics; Surrogate; Gaussian method; Turbulence

I. INTRODUCTION

Wind turbine optimization is primarily dependent on the wind velocity and wakes patterns within turbine blades. During the power extraction from the wind as the turbine blades move, a wind turbine experiences wake and turbulence effects, which results in the reduction of, wind velocity and ultimately power losses.

Different mathematical models were designed to examine and determine the turbine specifications which are implemented to estimate the output of wind turbines [1]. The pivotal factor for deploying different wind turbine optimization factors is wind velocity. Therefore, estimating wind velocity is vital for shaping the wind farm layout and power generation. Computational fluid dynamics (CFD) simulations were run over the rotating wind turbine blades to find the wind speed at various points. The turbine blades experienced turbulence and wake effects, which eventually reduced the wind speeds. This process took roughly eight hours to generate the results. To overcome this, the wind velocities at specific points were sampled and introduced into a machine-learning model. Different Surrogate models such as machine learning support vector regression, gradient boosting regression, and regression algorithms were formed. We resolved that these models can forecast wind speed inside wake and helped to calculate the upwind and downwind speeds. These models saved time for computationally expensive CFD and aero servo elastic turbine simulations. Moreover, these models will also help in assessing the fatigue in turbine blades and other failure factors. The wake effect through the wind turbine is shown in Fig. 1, the air velocity is reduced when it passes through the turbine blades.

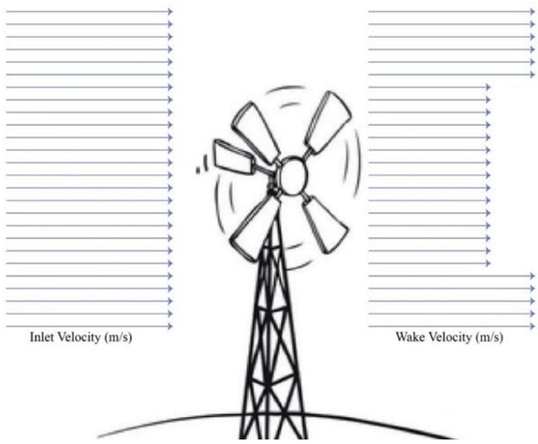

Fig. 1: Illustration of wake effect.

The deprivation of fossil fuels and the intensifying figures of global warming and environmental pollution has directed the development of green energy resources. Due to the depletion of hydrocarbon deposits world is facing an energy crisis. This situation has amplified the importance of power generation through green energy sources such as solar, hydropower, wind, and geothermal [2]. Among all the inexhaustible energy resources wind energy has shown remarkable advancement during the last ten years due to less environmental impact, cost efficiency, and sustainability [3]. Shakoor et al., [4] added that the appropriate site for the wind farm, improving the efficiency of wind turbines, reducing the wake effects within the moving turbine blades, and handling the technical issues are still a challenge for researchers and scientists. In brief, wind farm layout and wind turbine optimization are crucial to meet energy needs.

Designing, constructing, and maintenance of wind turbine farms cost a lot. Thus, the construction and position of turbines in a wind grange, wind velocity, and other dependent factors should be factored in before the installation of wind turbines. Several factors may alter the power obtained from wind turbines such as blade radius, air density, tower height, and size of the rotor. The most dominating factor is the variable wind, which generates wake and turbulence effects within rotating turbines blade and significantly decreases the wind velocity. Because of this, the power extracted from the turbine is reduced.

Researchers have introduced different mathematical models to minimize the wake effects such as Jensen's model, Larsen's model, and Frandsen's model. These models predicted that wake models' results were affected by the distance between wind turbines. These models cast-off heuristic techniques and old 2D methods that do not consider the rotation effects in wind turbine blades so their accuracy is not clear. There is a need for a modern method that can be hybridized for getting more optimal results [4]. CFD is a field within the realm of fluid mechanics that is practiced to solve complex engineering fluid problems by numerical means and analytical methods depicted that the exact solution to turbulent flows is impossible but the use of large-eddy simulations for optimizing wind energy has resolved the problem to a great extent. Based on Reynolds-Average Navier stocks equations (RANS) turbulence and wake effects can be calculated and improved. The standard turbulence closure schemes, k-ε, and k-ω models have been adopted to resolve the three-dimensional, steady-state (RANS) equations [5]. Stergiannis [6] further added that despite these models contributing a lot to velocity optimization, still lost wake data when moved to simplified models and they are computationally expensive. To minimize the computational workload for wake simulations, the blades are typically accounted for using the generalized actuator disk method, which represents the rotor through forces. To handle these forces correctly, a meticulous numerical approach is required. At present, there are three methods in practice, namely the actuator disk, actuator line, and actuator surface methods. To reduce the complexity of CFD simulations, the wind turbine blade model was generalized with the actuator disk approach. Three methods are used including the actuator surface model, actuator line, and actuator disk. The actuator disk model seems to be more efficient and widely used because of less computational effort. These models do not simulate near wakes because they are unable to capture tip vortices. These uncertainties originated from the discretization and turbulence modeling errors are still a challenge [7].

Steve Brunton [8] concluded that real-life optimization problems involve expensive computations and they are often applicable to a much smaller scale. In this regard, Machine Learning is guaranteed to offer a compatible and expeditious modeling outline that can be implemented to sort out various complications in fluid mechanics models, such as shape optimization, turbulence closure modeling, reduced-order modeling, control, and experimental data processing by using regression algorithms, Gaussian method, artificial neural network, and many others standard algorithms. These models require a substantial amount of data, which may not be available. Kim and Boukouvala [9] found that Surrogate modeling is one of the effective techniques to find the best-fit model or approximation that an algebraic model lacks. Implementing these techniques to wind turbine layout optimization will help to diminish the wake effect to great extent however, these models are unable to solve problems as they become complicated and enlarged dimensionally.

The surrogate modeling requires a large amount of data set that is usually not available. This paper aims to develop a map that will require a small data set to train the model. More notably, three different machine-learning techniques named linear regression, gradient boosting regression, and support vector regression were employed to find the best-fit technique for generating data in real-time, resolving high dimension problems, and calculating the error-prone in past methods.

II. The Geometry of the Blade and Fluid

In this paper, ANSYS Fluent and Workbench 2020 R1 were used to design the geometry of the blade and to perform simulations. This software used to be the best software to perform simulations and CFD calculations in industry and academia. ANSYS Fluent was used to run CFD calculations and to simulate the flow of fluid such as wind or air. To design the geometry of the blade and fluid, ANSYS DesignModeler was used. The geometry of the fluid was 120° or 1/3 part of a cone shape, with a length of 270 m and a single blade was enclosed inside the fluid geometry. The fluid (air) flows throw the cone from the inlet and upper inlet, the inlet was placed 90 m away from the blade at (0, 0, 90) and the blade was placed at the center at (0, 0, 0). The blade geometry is revealed in Fig. 2.

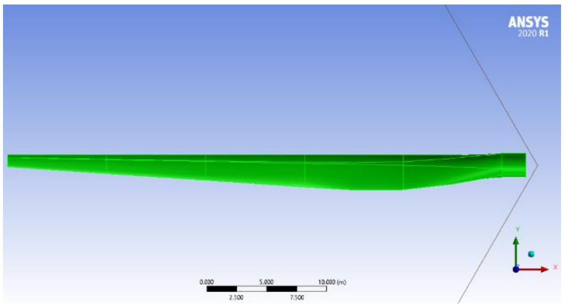

Fig. 2: Blade geometry with ANSYS Fluent 2021.

A. Meshing

After completing the geometry of the blade and fluid region the shape of the mesh must be finalized. Mesh is a process in which flow domains split into much smaller subdomains. These mesh subdomains can be 2D or 3D in shape. In this paper, simple 3D shapes were used. The fluid flow is governed by partial differential equations, and it is almost impossible to solve these equations analytically, consequently, it is crucial to split the flow domains. The governing equations are solved inside each subdomain. The precision of the outcomes and runtime is determined by the number of these subdomains, more subdomains mean solving more equations and longer runtime, and more accurate results. The mesh around the blade was much finer to get a more accurate effect on the wind.

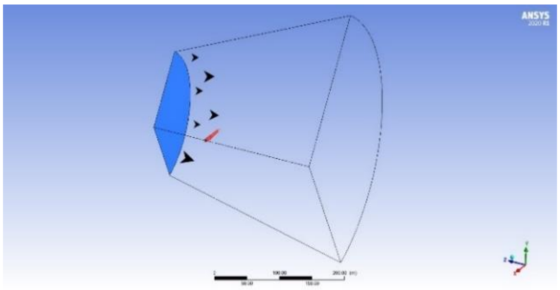

Fig. 3: Turbine blade model and fluid region geometry description.

The wind flows in red arrows in the negative z-direction from the inlet in blue and the blade is colored in red. The length of the cone is 270 m along the negative z-axis. The figure displays only a small subset of the entire simulation, which encompasses a significantly larger number of data points.

B. Experimental setting

To set up the parameters and type of the experiment ANSYS Fluent was used to run the experiment. A total of eight experiments were conducted at varying wind velocities ranging from 5 m/s to 19 m/s. Check Tables 1 and 4 for a complete list of wind velocities used. In ANSYS Fluent SST k-ω model was used as the viscous model. The boundary conditions were defined, and the blade was characterized as a solid surface. Wall is a surface to which no fluid can pass through it. To induce blade rotation, a frame motion was implemented, with the angular velocity set to a maximum value based on experimental data, as defined in [10]. For at 12 m/s speed of the fluid, the rotational speed of the blade was computed as 1.164. A complete list of wind velocities and rotational velocities is shown in Table 1.

The fluid flows throw the inlet and the upper inlet in a negative z-direction, and the initial velocity of the fluid was set at one of the described velocities. After that, the solution method was selected, which defines how the experiment is going to be performed and iterated. Here the number of iterations was set to 1000 and then initiated.

$$\omega = \frac{2\pi v}{n s} \quad (1)$$

$$TSR = \frac{U}{v} \quad (2)$$

The formulas for ω and TSR are provided below, with v representing the wind speed, n indicating the number of blades representing the length of the disturbed air stream, and U representing the blade's angular velocity.

C. Sampling experiment results

After running and completing the calculations of the experiment, ANSYS CFD-Post was used to get the results. The values changing throughout the experiment can be exported.

Table 1: Input and output variable of CFD model.

No. of Observations	Input Wind Velocity (m/s)	Angular Velocity (rad/s)	Tip Speed Ratio (TSR)	Output Inside Wake Wind Velocity ω/ (rad/s)
1	6	0.582	4.014	5.92
2	9	0.873	4.014	8.91
3	12	1.164	4.014	11.89
4	15	1.164	3.212	14.9
5	18	1.164	2.676	17.9

Table 2: Input and output variables to validate the machine learning models.

No. of Observations	Input Wind Velocity (m/s)	Angular Velocity (rad/s)	Tip Speed Ratio (TSR)	Output Inside Wake Wind Velocity ω/ (rad/s)
1	8	0.776	4.014	7.9
2	14	1.164	3.441	13.99
3	20	1.164	2.408	11.9

In this experiment, a streamline was used to get the velocities. To get the absolute velocities at certain points in the wake, velocity in the stationary frame was exported and the sampling method was set to equally spaced and the number of sampling gradient points was set to 100,000. The average velocity in the stn frame was calculated to get a single value. Tables 1 and 2 give the calculated values. This process was repeated for each value of wind velocity.

III. SURROGATE MODELS OF MACHINE LEARNING

Wind-power plant optimization depends on the air input velocity and arrangement of turbines. The wake produced by the first turbine reduces the input velocity for the next turbine. Velocity inside the wake ω is calculated by performing simulations of CFD [11]. However, it takes approximately 8 hours to calculate a single reading in the CFD model, depending upon the strength of the computer. Therefore, substitute models (Fig.4) are used to get output in real-time by using techniques of machine learning [12] are exposed in Table 3.

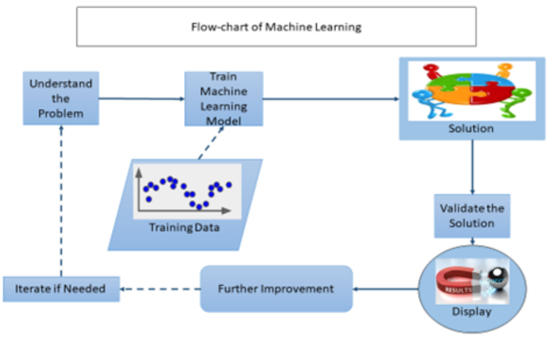

Fig. 4: Brief visualization of machine learning working process which explanes each and every step in detail by visualizing it as flow diagram.

Machine learning models are typically trained using a dataset that is split into two parts: the training set and the testing set. The training set is used to train the model on how to make predictions or decisions, while the testing set is used to evaluate the accuracy of the model. The accuracy of the model depends on the quality and quantity of the training data used. Larger datasets with more diverse data points can lead to improved accuracy.
It is important to ensure that the training data used to train the model is representative of the population that the model is intended to predict for. Failure to do so may result in inaccurate predictions when the model is applied to new data. Hence, selecting the right features and data points for the training set is crucial for the model's performance. To summarize, the quality and representativeness of the training data are essential for the accuracy of machine learning models, which rely on sample data points for both training and testing.

A. *Training of surrogate models*

Substitute models were trained on the data set gathered from CFD model simulations Table 1. Three types of models were trained; Support Vector Machine Regression, Linear Regression, and Gradient Boosting Regression. Besides the training of substitute models, different types of errors were also computed to check and compare the model's Table 4. All the models trained well and showed acceptable errors in prediction revealed in Table 7. The comparison of errors of each model concluded that gradient boosting regression is the best among them for further predictions, as shown in Fig 5.
Table 3: Machine Learning Substitute Models used for Experimentation in Jupyter (anaconda 3.7).

Machine Learning Algorithm	Description
LR [13]	Linear Regression describes the relationship between two variables and predicts the outcome depending on the training data set.
SVMR [14]	Support Vector Machine Regression is an example of a supervised machine learning algorithm that is commonly employed for making predictions of continuous values.
GBR [15]	Gradient Boosting Regression is the best algorithm for the prediction of our data with the least root mean square error.

Governing equation of linear regression:

$$\hat{y}=\theta_0+\theta_1 x_1+\theta_2 x_2+\cdots+\theta_n x_n \quad (3)$$

$$\text{RMSE}(X, h)=\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right)-y^{(i)}\right)^2} \quad (4)$$

The predicted value is represented by $\hat{y}$, and the model's constraint vector θ contains the bias term θ_0 and the feature weights θ_1 to θn. The illustration feature vector x contains x_0 to x_n, with x_0 always set to 1. The dataset contains m values, and x(i) characterizes the function of input while y(i) represents the function of output. The prediction function is represented by hypothesis h.

Table 4: Description of errors calculated.

Errors	**Description**
MAE	Mean Absolute Error
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Biased Error
RS	Root Square
RMSE	Root Mean Squared Error

B. Validity and Accuracy of Surrogate Models

All the substitute models of Machine Learning were trained on the data obtained from the CFD model run in Jupyter Notebook (anaconda 3) [16,17] and their graphs clarified that gradient boosting regression is the best to predict Fig. 3.

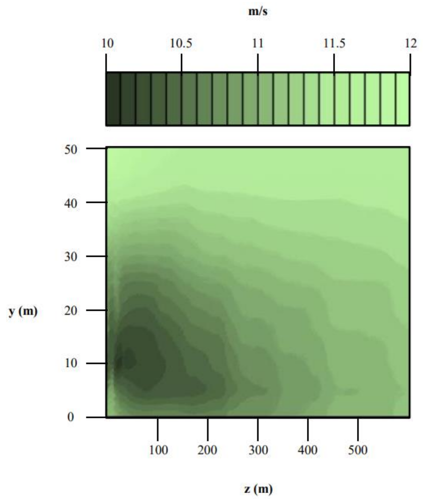

Fig. 5: The gradient boosting regression model was trained on all data sets. Scales of both axes were fixed in meters. While the range of axes is fixed as y ϵ [0, 50] were z ϵ [0,600]. Detailed iterations were made for more accuracy. This figure just demonstrates a small slice of the conical wake region.

The performance of all models was verified on a data set of velocities of 6, 9, 12, 15, and 18 m/s which is shown in Table 1. Initially, the models were trained on the complete data set of the CFD model. Then, another value was calculated from the CFD model to crosscheck the prediction of the model Table 6. The GBR model predicted the value with an absolute mean error of 0.000096.

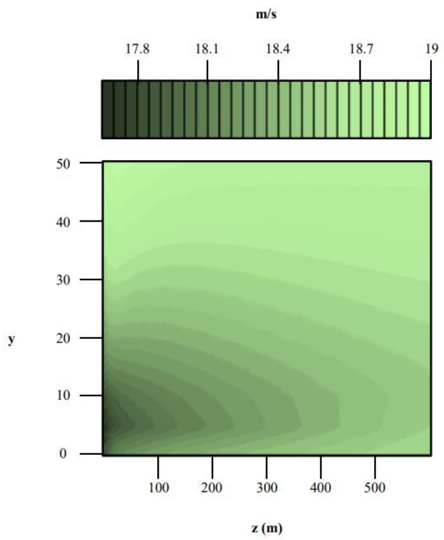

Fig. 6: This figure shows the prediction of support vector machine regression on wind velocity of 17.5 m/s after being trained on all data set points.

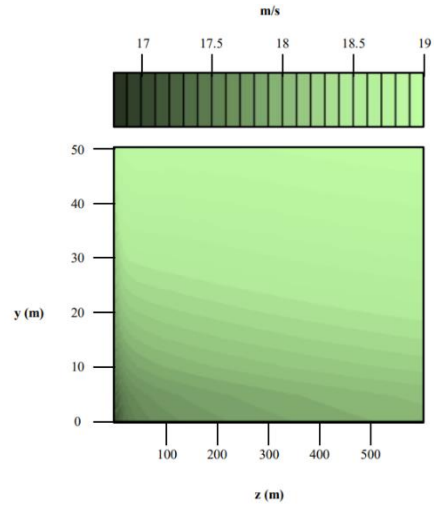

Fig. 7: This figure demonstrates the prediction of linear regression on wind velocity of 18.1 m/s after being trained on all data set points.

As the prediction error is acceptable, there is no need to waste 8 hours for a single value in the CFD model. Nevertheless, Gradient Boosting Regression is the best machine learning substitute model [18]; it can be perused for further results in Table 5 which also prediction of support vector machine regression can be seen in Fig 6.

Table 5: Comparing selected input and output values of CFD model with all three machine learning models;LRM, SVMR and GBR to check the validity.

No. of Observations	Input Air Velocity (m/s)	Output Air Velocity Inside the wake (m/s)			
		CFD Model	Linear Regression Model	Support Vector Machine Regression Model	Gradient Boosting Regression Model
1	6	5.92	5.921667	6.019747	5.92015
2	9	8.91	8.906667	8.904747	8.91008
3	12	11.89	11.89166	11.78974	11.8900
4	15	14.90	14.90000	15.00050	14.8999
5	18	17.90	17.90000	17.53997	17.8998

Table 6: Here is the reverse engineering techn ique to validate the results of machine learning models. First prediction was made by models then it was calculated by CFD model to validate that how much the models are reliable for randome inputs.

Input	Output			
	CFD Model	Linear Regression Model	CFD Model	Gradient Boosting Regression Model
5.5	5.43	5.472570	5.512458	5.493258

Table 7: Comparing the errors of all three machine learning models to find out which one is more reliable.

Models	Errors					
	RSE	MBE	MAE	MSE	RMSE	MAPE
Linear Regression	1	0	0.001333	3.33E-06	0.001826	0.000159
Support Vector Regression	0.998219	0.053056	0.133157	3.19E-02	0.178745	0.010546
Gradient Boosting Regression	1	0	0.000096	1.27E-08	0.000113	0.00001

IV. CONCLUSION

Wind turbine power optimization has been studied through different strategies involving different parameters and assumptions. The overarching variable in turbine efficiency is wind velocity which is studied and analyzed in this paper. The results showed that the amalgamation of machine learning and CFD produced the desired results with better accuracy to predict wind speed. Among all the three models; Linear Regression, Support Vector Regression and Gradient Boosting Regression, the last on was the most reliable model as it predicted the out put of 5.920159 which is more accurate to actual out of 5.92 calculated by CFD model. Moreover, Gradient Boosting Regression has also proved to be best with respect to error parameters as well with the least Mean Baised Error of 0.

Models were trained on linear variables like 6, 9 12 etc., which is easy to train and takes minimum time to be ready for prediction to give the best possible predictions. However, if models were trained on non-linear variables like x^2, y^2, and z^2 and their inverses instead of the linear ones; x, y, and z, the models would take much more time to be trained. Moreover, it would be too complex to formulate but it will minimize the error ratio and gives the best prediction. There are even many other best ways to formulate the substitute models such as modeling on polynomial regression, logistic regression, or formulating decision trees, random forest, and naive bayes. Further, clustering as well as combining the models to do their best in prediction with minimized error. All of the above-discussed techniques belong to Machine Learning. Besides them, a subset of Machine learning called Deep Learning and ANN (Artificial Neural Network) can also be used.

Nomenclature:

U : Angular velocity of the blade
v : Wind speed
n : Number of blades
s : Length of the disturbed air stream
x(i) : Input function
y(i) : Output function
h : Prediction function
$\hat{y}$: Prediction value
θ : Model's parameter vector

REFERENCES

[1] Samorani, M., The wind farm layout optimization problem, in Handbook of wind power systems. (2013), Springer. p. 21-38.

[2] Dang, T. Introduction, history, and theory of wind power. in 41st North American Power Symposium. (2009). IEEE.

[3] Tong, W., Fundamentals of wind energy. Wind power generation and wind turbine design, (2010): p. 3-48.

[4] Shakoor, R., et al., Wake effect modeling: A review of wind farm layout optimization using Jensen′ s model. Renewable and Sustainable Energy Reviews, (2016). 58: p. 1048-1059.

[5] Sumner, J., C.S. Watters, and C. Masson, CFD in wind energy: the virtual, multiscale wind tunnel. Energies, 3(5) (2010): p. 989-1013.

[6] Stergiannis, N., et al. CFD modelling approaches against single wind turbine wake measurements using RANS. in Journal of Physics: Conference Series. (2016). IOP Publishing.

[7] Sanderse, B., S. Van der Pijl, and B. Koren, Review of computational fluid dynamics for wind turbine wake aerodynamics. Wind energy, 14(7) (2011) : p. 799-819.

[8] Brunton, S.L., B.R. Noack, and P. Koumoutsakos, Ma-chine learning for fluid mechanics. Annual Review of Fluid Mechanics, 52 (2020) : p. 477-508.

[9] Kim, S.H. and F. Boukouvala, Machine learning-based surrogate modeling for data-driven optimization: a comparison of subset selection for regression techniques. Optimization Letters, 14(4) (2020) : p. 989-1010.

[10] Ragheb, M. and A.M. Ragheb, Wind turbines theory-the betz equation and optimal rotor tip speed ratio. Fundamental and advanced topics in wind power, 1(1) (2011): p. 19-38.

[11] Kuo, J.Y., et al., Wind farm layout optimization on complex terrains–Integrating a CFD wake model with mixed-integer programming. Applied Energy, (2016). 178: p. 404-414.

[12] Andersson, L.E. and L. Imsland, Real-time optimization of wind farms using modifier adaptation and machine learning. Wind Energy Science, 5(3), (2020): p. 885-896.

[13] Seber, G.A. and A.J. Lee, Linear regression analysis. Vol. 329. (2012): John Wiley & Sons.

[14] Awad, M. and R. Khanna, Support vector regression, in Efficient learning machines. 2015, Springer. p. 67-80.

[15] Zhang, Y. and A. Haghani, A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies, (2015), 58: p. 308-324.

[16] Otchere, D.A., et al., Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. Journal of Petroleum Science and Engineering, (2021), p. 109244.

[17] Zhou, Z. H. (2021). Machine learning. Springer Nature.

[18] Bell, J. (2022). What is machine learning?. Machine Learning and the City: Applications in Architecture and Urban Design, 207-216.