

Received January 11, 2022, accepted January 25, 2022, date of publication January 28, 2022, date of current version February 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3147602

# Machine Learning-Based Digital Twin for Predictive Modeling in Wind Turbines

MUHAMMAD FAHIM<sup>1</sup>, VISHAL SHARMA<sup>1</sup>, (Senior Member, IEEE), TUAN-VU CAO<sup>2</sup>, BERK CANBERK<sup>3</sup>, (Senior Member, IEEE), AND TRUNG Q. DUONG<sup>1</sup>, (Fellow, IEEE)

<sup>1</sup>Queen's University Belfast, Belfast BT7 1NN, U.K.

<sup>2</sup>Norwegian Institute for Air Research, 2027 Kjeller, Norway

<sup>3</sup>Istanbul Technical University, 34469 Istanbul, Turkey

Corresponding author: Trung Q. Duong (trung.q.duong@qub.ac.uk)

This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P019374/1, and in part by the EPSRC Impact Acceleration Accounts (IAA) Award.

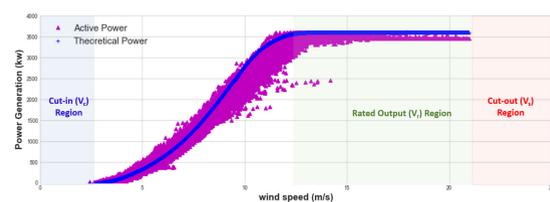
**ABSTRACT** Wind turbines are one of the primary sources of renewable energy, which leads to a sustainable and efficient energy solution. It does not release any carbon emissions to pollute our planet. The wind farms monitoring and power generation prediction is a complex problem due to the unpredictability of wind speed. Consequently, it limits the decision power of the management team to plan the energy consumption in an effective way. Our proposed model solves this challenge by utilizing a 5G-Next Generation-Radio Access Network (5G-NG-RAN) assisted cloud-based digital twins' framework to virtually monitor wind turbines and form a predictive model to forecast wind speed and predict the generated power. The developed model is based on Microsoft Azure digital twins infrastructure as a 5-dimensional digital twins platform. The predictive modeling is based on a deep learning approach, temporal convolution network (TCN) followed by a non-parametric k-nearest neighbor (kNN) regression. Predictive modeling has two components. First, it processes the univariate time series data of wind to predict its speed. Secondly, it estimates the power generation for each quarter of the year ranges from one week to a whole month (i.e., medium-term prediction) To evaluate the framework the experiments are performed on onshore wind turbines publicly available datasets. The obtained results confirm the applicability of the proposed framework. Furthermore, the comparative analysis with the existing classical prediction models shows that our designed approach obtained better results. The model can assist the management team to monitor the wind farms remotely as well as estimate the power generation in advance.

**INDEX TERMS** Digital twin, machine learning, predictive models, wind energy, 5G.

## I. INTRODUCTION

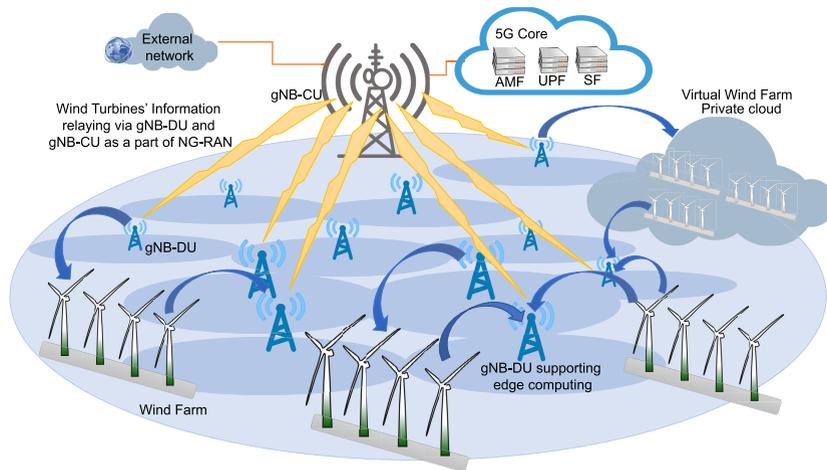
Nowadays, wind farms are a common sight in the United Kingdom to generate clean energy as well as contribute to achieving the net-zero emission goal by 2050 [1]. One of the primary sources of harmful emission is the generation of electricity using fossil fuels (i.e., coal and gas). To keep the environment clean, wind energy is a plentiful source to generate environment-friendly energy. The wind turbines allow us to use the power of the wind to rotate generators and produce energy. The wind turbine performance varies with the change of season as well as geographic location [2], [3]. Therefore, the same wind turbine has different performances in different months and at various locations. It poses challenges for the

The associate editor coordinating the review of this manuscript and approving it for publication was Mounim A. El Yacoubi .



**FIGURE 1.** An active power generation and theoretical power curves concerning wind speed are presented. A theoretical power curve provides information about the differences in generated power from the wind turbine as well as with theoretical power.

energy management team to handle uncertainties. To monitor the performance, the modern wind turbines are equipped with supervisory control and data acquisition (SCADA) unit [4]. It is a cost-effective and data-driven approach to store the data streams for further analysis. The selected data dimensions from the SCADA unit are presented in the following Fig. 1.



**FIGURE 2.** An illustration of 5G network-assisted architecture for digital twin to forecast the generated power by wind turbines.

In Fig. 1, the vertical axis presents the power generation with its units kilowatt (kW), and the horizontal axis presents wind speed with its unit meter per second (m/s). It provides the theoretical power and active power curve with cut-in ( $V_c$ ), rated output ( $V_r$ ), and cut-out ( $V_s$ ) region. The ( $V_c$ ) region specifies where the wind has no power to generate the energy. The region between ( $V_c$ ) and ( $V_r$ ) is a rapid growth power generation that depends on wind speed. Similarly, ( $V_r$ ) is obtained rated output of the wind turbine, and ( $V_s$ ) region shows the power generation stop due to high winds. The power curves present the non-linear relationship between the generated power and the wind speed incidence at the height of the rotor hub [5]. The dynamical behavior can be expressed as:

$$P_w(v) = \frac{1}{2} \rho A v^3 \quad (1)$$

where  $P_w$  is power associate with wind speed  $v$  of wind,  $\rho$  refers to air density, and  $A$  is the surface area of the turbine rotor.

The data-driven approaches and curve fitting techniques are developed to predict the performance [6], condition monitoring [7], and fault diagnostics [8] of wind farms. At the same time, the current challenges include accurate forecasting of power generation and the limitation of real-time data exploration of the whole wind farm. The developed forecasting approaches are based on statistical [9] and machine learning models [10], [11]. Yang *et al.* [12] introduced a data preprocessing technique to estimate the data distribution over the quantile of years. They proposed a statistical approach based on bi-directional Markov chain interpolation for theoretical power calculation. It helps to obtain realistic results. Yun *et al.* [2] developed a statistical framework based on multiple spline regression model for the power curve modeling of the wind turbine. Their approach described the complex nonlinear relationship between wind speed and wind power using different basis functions and

different numbers of knots inside multiple spline regression. Khosravi *et al.* [11] applied machine learning algorithms to predict the wind speed for Osorio wind farm in the south of Brazil. They applied neural networks, support vector regression, and fuzzy inference systems optimized with computational intelligence-based algorithms. They reported a neural network-based model outperforms as compared to the considered model in the study. Furthermore, they conclude wind speed has a direct influence on the generated power. In our previous work [6], we also conclude the relationship between wind and generate power followed by deep learning model prediction.

The recent advancement in machine learning approaches especially deep learning models has a breakthrough to solve complex problems. In case of temporal sequence modeling, deep learning approaches has the ability to consider the long history of the input data which leads to an accurate result as compared to its classical approaches. Secondly, a huge amount of data is available from SCADA units that make such models more robust to learn and generalize the concept of prediction. In the continuation of our research towards wind energy, the developed model considers wind speed as univariate time-series data to forecast the whole month and each week. Later, this wind speed forecasting is further used to predict power generation.

In terms of networked systems, digital twins have been considered as an emerging technology to enable a wide range of typical applications such as manufacturing, 5G and beyond networks, intelligent transportation systems, climate change, and smart cities [13]. In [14], a digital twin-based intelligent cooperation framework of UAV swarm integrated with machine learning algorithm is investigated to tackle the problem of real-time control of the behaviours of UAV swarm. Digital twins can additionally evolve their context-awareness capabilities to identify cybersecurity issues in real-time, which can be effectively applied to smart grid deployments [15]. More importantly, digital

twins technology is a powerful tool to address problems of joint communication and computation task offloading in mobile edge computing (MEC) [16] to enable various mission-critical applications in the industrial Internet of things [17], [18]. Therefore, designing digital twins-based solutions significantly contributes to the development of both academia and industry in the digital era.

The real-time data exploration and a feedback loop to the wind farms are possible through digital twins' technology. It provides next-generation computer-oriented solutions. It can create a digital copy of wind farms connected with the physical wind turbines, where supervisory control and data streams are accessible for analysis and prediction. Olatunji *et al.* [19] briefly introduced the digital twin technology in wind turbine fault diagnosis and condition monitoring. They highlight the wind industry transformation to the next level with enhanced accessibility and availability using digital twins technology. Similarly, Kishnamoorthi *et al.* [20] developed a digital twin based model to predict the remaining life of offshore fixed and floating wind turbines as a predictive maintenance strategy. The model is based on operational data of SCADA units and a physics-based approach to predict the remaining life of wind turbines.

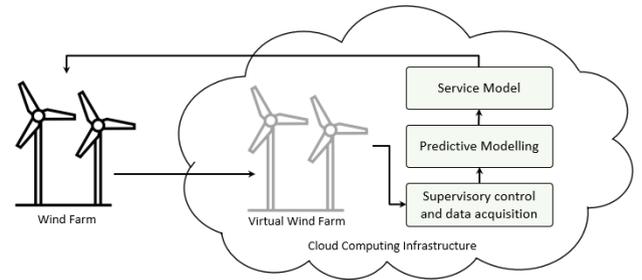
Many researchers have designed various digital twin models for predictive maintenance and have contributed significantly to improvements in wind energy technology. At present, none of the existing digital twins can predict power generation and real-time monitoring. Therefore, this study proposes a novel framework to process the temporal data stream to forecast the wind speed and predict the generated energy. Consequently, it can provide virtual access to the wind turbines 24 by 7 without visiting the physical wind farms. While its feedback loop makes the communication back and forth with wind turbines for monitoring purposes. The contribution of our work is outlined as follows:

- Our framework is built over a 5G-Next Generation-Radio Access Network (5G-NG-RAN) assisted cloud-based digital twin model for understanding and analyzing wind farms. It is a cost-effective solution; hence, digital twins modeling is possible with the pay-as-you-go cloud services.
- The designed machine learning pipeline has two novel components—first, the forecasting of wind speed based on an advanced temporal convolutional neural (TCN) network. Second, the processing of wind forecast to predict the power generation for a month, including each quarter (i.e., medium-term analysis).

The rest of this paper is organized as follows. Section II presents the proposed framework details with methods and procedures. Section III, provide the results, discussion, and performance of the framework. Finally, Section IV concludes our findings.

## II. NETWORK MODEL, METHODS AND PROCEDURES

The network-assisted prediction allows support from edge infrastructure to form a setup of a collaborative system



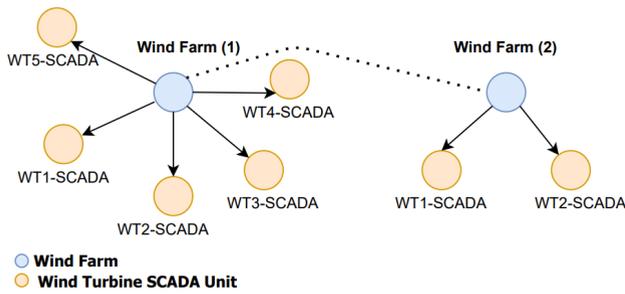
**FIGURE 3.** A cloud-based framework to model the wind farm using digital twins technology and predictive modeling to forecast the generated power.

between the physical wind farm and the digital twin of it. We specifically considered Next Generation Radio Access Network (NG-RAN) as a core architecture shown in Fig. 2. The NG-RAN divides gNB into a Control Unit (CU) and a Distributed Unit (DU). gNB-CU forms the core of the network handling 5G functions such as Access and Mobility Management Function (AMF), User Plane Function (UPF) and all the associated Security Functions (SF) [21]. The gNB-DU forms the edge part of the network interacts with the wind farm through gateways and relays information via gNB-CU to the private cloud setup with a virtual (digital twin) wind farm to perform predictions. The advantage of using 5G-NG-RAN is that it allows better integration of cloud services as wind farms are geographically isolated regions than the data centers and can also ensure better coverage with lower latency offering more real-time services. It also reduces the cost of deployment by reducing the number of near-farm data centers. Moreover, better resource utilization is attainable if a large amount of data from the wind turbines must be shared between the physical and the digital systems. Furthermore, a basic framework considering the interaction between the wind farm and the virtual farm without the network infrastructure, illustrated in Fig. 3, helps to understand the actual workflow. It offers a real wind farm that is connected to its digital twin “virtual wind farm”. Each wind turbine has supervisory control and data acquisition unit in a wind farm to provide the data for monitoring. The captured data logs are connected with virtual wind turbines with the help of digital twins modeling. Our framework processed the data logs to perform predictive modeling. It has the ability to report the possible generation of electric power from the wind farm in the coming days.

### A. DIGITAL TWINS MODELING

The wind turbine supervisory control and data acquisition (SCADA) unit is modeled as digital twins using the Microsoft Azure platform. It is based on platform as a service (PaaS) to model digital twins and provides digital monitoring as a next-generation computer-oriented solution. Furthermore, a cloud-based infrastructure provides a cost-effective solution. The 5D modeling approach [22] is followed to model the wind farms. The Eq. 2 depicts the model.

$$\Omega_{DT} = (PE, VR, DC, CS, Ss) \quad (2)$$



**FIGURE 4.** A knowledge graph of wind farms which are connected to monitor the behavior of each wind turbine.

where  $\Omega_{DT}$  is digital twins wind farm,  $PE$  is physical entities,  $VR$  is virtual representation,  $DC$  is data curation,  $CS$  is communication scheme, and  $Ss$  for services. The details about each dimension is presented as follows:

1) PHYSICAL ENTITIES (PE)

The wind turbines consist of various physical entities, including mechanical devices, monitoring sensors, SCADA units, and activities processes. The  $PE$  can be categorized into a unit level, system level, and system of system-level [23]. The SCADA unit is considered at  $PE$  level, which is an essential tool to collect the data for monitoring the behavior of wind turbines. It enables the analyst to access the historical and real-time data of wind turbines for further analysis.

2) VIRTUAL REPRESENTATION (VR)

An entity in the virtual environment represents each SCADA unit of a wind turbine to construct the wind farm. The  $VR$  has the ability to build a connection with the wind turbines and replicate the behavior virtually. The wind farms can be connected even they are physically apart using a knowledge graph as shown in Fig. 4.

In Fig. 4, two wind farms are connected using a knowledge graph with five and two wind turbines SCADA units, respectively. Similarly, each wind farm has an interface that connects the SCADA unit of wind turbines. A relationship and association are defined using digital twins definition language (DTDLD) to join the SCADA unit. The DTDLD is an XML-based language, and its snapshot is present in appendix Fig. 13. It also has the ability to provide a mechanism to connect sensors and link the real-time readings for further analysis.

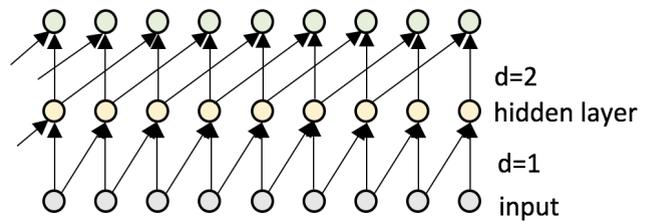
3) DATA CURATION (DC)

The  $DC$  is the central part of digit twins. The temporal data streams of SCADA units are curated to monitor the SCADA units. It provides real-time access to the  $PE$ . The following output presents the wind speed, direction, generated power, and theoretical power of the wind turbines. Furthermore, it also gives metadata information.

The connected SCADA unit presents the real-time values in the digital twin explorer of Microsoft Azure. It shows the

```
object {7}
  $dtId F2WT-SCADA
  $etag : W/"d091eef3-971e-4467-989d-3c54259e3deb\"
  windSpeed : 5.3
  windDirection : 259.9
  generatedPower : 380.04
  theoreticalPower : 416.32
  $metadata {5}
    $model {1} : dtmi:windfarm:Windturbine;1
    windSpeed {1}
      lastUpdateTime : 2021-09-02T14:22:11.4599211Z
    windDirection {1}
      lastUpdateTime : 2021-09-02T14:22:11.4599211Z
    generatedPower {1}
      lastUpdateTime : 2021-09-02T14:22:11.4599211Z
    theoreticalPower {1}
      lastUpdateTime : 2021-09-02T14:22:11.4599211Z
```

**FIGURE 5.** The connected SCADA units data in Microsoft Azure digital twin explorer.



**FIGURE 6.** Causal convolution block with two layers with kernel size 2.

digital twin’s ID, etag, real-time values of the SCADA unit, and metadata information about the last update of the specific unit. It also has powerful structure query language support to have deeper analysis inside the constructed knowledge graph. A service model could be deployed over it to make an informed decision.

4) COMMUNICATION SCHEME (CS)

Digital twins are dynamically connected with  $PE$  units using a representational state transfer application programming interface (REST-API). The  $CS$  also connects the  $DC$  for real-time communication. Such  $CS$  enables the functionality of the digital twins model to communicate in real-time. It is also responsible for data flow from the  $PE$ -SCADA unit to  $VR$ -SCADA units.

5) SERVICES (Ss)

The  $Ss$  is an essential part that provides an adapter to communicate with other  $PE$ , model services, data analysts, etc. It provides support to customized services that can be built outside of the model. The proposed framework is based on cloud computing infrastructure, which already has a standards protocol for service delivery.

**B. PREDICTIVE MODELING**

The predictive modeling component process two sources of information. First, a wind speed as a univariate time series using a novel deep learning model based on a temporal convolutional neural (TCN) network. The TCN is modern deep neural architecture and has proved better results as

compared to its sequence modeling counterparts as well as more efficient in terms of computation time [24]. It combines the power of dilated convolution and residual block. Second, the forecasting results are further processed to predict the power generation from wind turbines using k-nearest neighbor regression. The input to the TCN is wind speed data stream, and it is defined as:

$$WS = (ws(t_1), ws(t_2), ws(t_3), \dots, ws(t_n))^T \quad (3)$$

where  $WS$  is wind speed,  $t_{i+1} = t_i + \Delta t$  and  $WS = (\cdot)^T$  represents the value of wind speed at any time instance  $t$ . The basic assumption is made that  $p(ws_{t+1}|ws(t_1), ws(t_2), ws(t_3), \dots, ws(t_n))$  does not depend on the future timestamps. It is a realistic assumption in case of wind speed forecasting. A function  $f$  is defined as:

$$f(WS) \mapsto ws(t_{n+m}) \quad (4)$$

where  $ws(t_{n+m})$  denotes the forecasting of month or any quarter inside it. The function  $f$  contains the dilated convolution layers residual blocks. The dilated convolution is defined as:

$$C(ws) = \sum_{i=0}^{k-1} f(i)ws_{h-d.i} \quad (5)$$

where  $C$  is dilated convolution operation,  $k$  is the filter size being learned, and  $h-d.i$  consider the past sequence of wind speed. It allows the network to operate on a coarser scale rather than a normal convolution but more efficient as shown in Fig. 6.

The following expression computes the layer of TCN:

$$\text{Number of layers} = \left\lceil \frac{(\text{input length} - 1)}{\text{kernel} - 1} \right\rceil \quad (6)$$

A mean squared error (MSE) loss function is used for training to converge the temporal convolution neural network. It is calculated as:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

The output is passed to machine learning model k nearest neighbor (kNN) regression [25] to regress the value of energy. It is a supervised nonparametric regression technique that calculates the distance of the test point in the feature space. The goal is to predict the power generation as a linear combination of its  $k$  nearest neighbors using a distance metric. The  $k$  is a hyper-parameter that indicates the number of neighbors to be considered for the prediction of  $GP$ . We find out the optimal number of neighbors (i.e.,  $k = 9$ ) using grid search. Its graph with further discussion is presented in Section III. The input to the model is predicted wind speed  $WS$  that needs to predict the generated power  $GP$ . It can be defined as a function:

$$f(WS) \mapsto GP \quad (8)$$

It enables the prediction of generated power over the forecasting of wind speed. Consequently, it can assist the team makes informed decisions over the predicted energy.

TABLE 1. The wind turbine with data collection information.

Wind turbine	Dataset information
Manufacture	Sinovel wind group company
Wind turbine capacity	1.5 MW (Mega Watt)
Dataset start data	January 1, 2018
Dataset end date	December 31, 2018
Data frequency	10 minutes
Data validation	SCADA operator

### C. DATASET

A publicly available onshore wind farm dataset [26] is used for our experiment. The wind farm is located in Yalova, the northwestern region of Turkey, and operation since 2016. About the regional information, Yalova is located at the latitude of 40.65502 N and the longitude of 29.27693 E in the Marmara region. The deployed wind turbines are manufactured by Sinovel wind group company. These wind turbines are equipped with a SCADA unit, and four measurements are reported at 10 minutes intervals. (1) *Active power*: it reports the power generated by the wind turbine using the stochastic nature of wind speed. (2) *Wind speed*: it measures the speed at the height of the rotor hub of the wind turbine. (3) *Wind direction*: it provided the direction of the blades that turns automatically to the direction where the wind blows. (4) *Theoretical power*: it is the power value computed by the manufacturing formula inside the control system of the SCADA unit. Our experiments are based on the data recorded during the period of January 1, 2018 – December 31, 2018. The data was collected and stored by SCADA unit. The SCADA units have the ability to store, retrieve, and exports the data for a variety of stakeholders. The SCADA system operator has a responsibility to validate the incoming data. The wind turbine with data collection information is present in Table. 1.

The active power generation of a wind turbine has a strong correlation with wind speed. The active power is widely employed for monitoring wind turbine performance and power generation [27], [28]. We considered the wind speed as a temporal data series to predict the intensity of the wind in the coming days. Based on wind prediction we further forecast the “active power” generation. The obtained results are explained in the following section.

## III. RESULTS

This section explains and presents the obtained result, followed by a discussion.

### A. WIND FORECASTING

To understand the data, the data is explored. We found that the dataset contains missing values at a few time intervals that indicate maintenance of wind turbines or other possible reasons. We replace such value with previous time-stamp observation. It is a necessary step to process the data in a

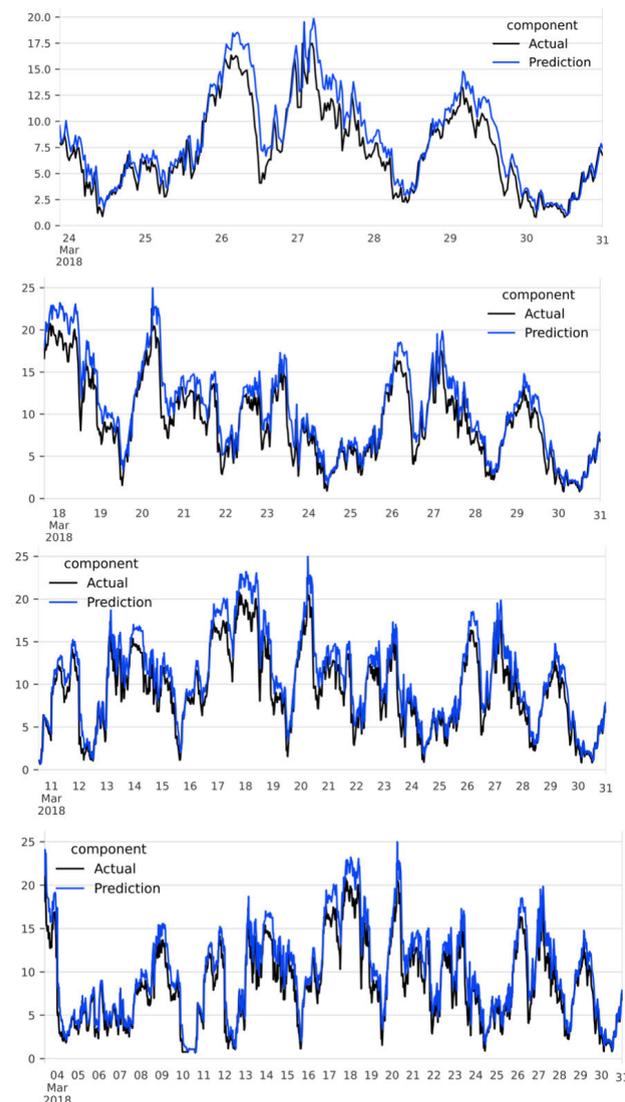


FIGURE 7. First quarter wind prediction for 7, 14, 21 and full month.

time-series manner. We split the dataset into four quarters and make predictions for a week, two weeks, three weeks, and the entire month of each quarter  $Q1, Q2, Q3,$  and  $Q4$ . The training model parameters are presented in Table 2

TABLE 2. The TCN model hyperparameters.

Hyperparameters	Value
Number of epocs	50
Dialation	2
Kernel size	5
Number of filters	6
Dropout	0.1

The following two performance metrics are considered to measure the performance of the model.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

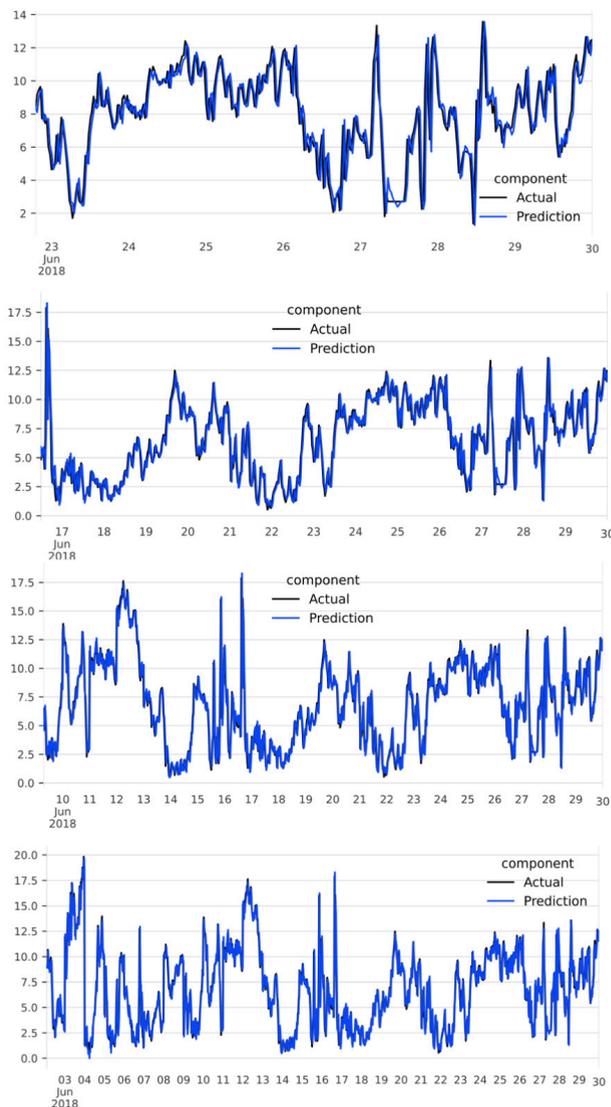


FIGURE 8. Second quarter prediction for 7, 14, 21 and full month.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

where  $MAE$  is mean absolute error,  $RMSE$  is the root mean squared error,  $y$  is the actual value from the test set, and  $\hat{y}$  is the predicted value from the trained model. The  $MAE$  measures the average magnitude of the errors by considering the absolute value, which presents the accuracy of the prediction. The  $RMSE$  measures the forecasting error by differencing the prediction and the actual value, which is squared, average, and then followed by a square root. Both  $MAE$  and  $RMSE$  can be used together to present the model errors. The  $RMSE$  provides large error values as compared to  $MAE$  because it gives relatively high weight to large errors. If both performance measures have the same value then it means the error of the model has the same magnitude. The first quarter prediction is presented in Fig. 7.

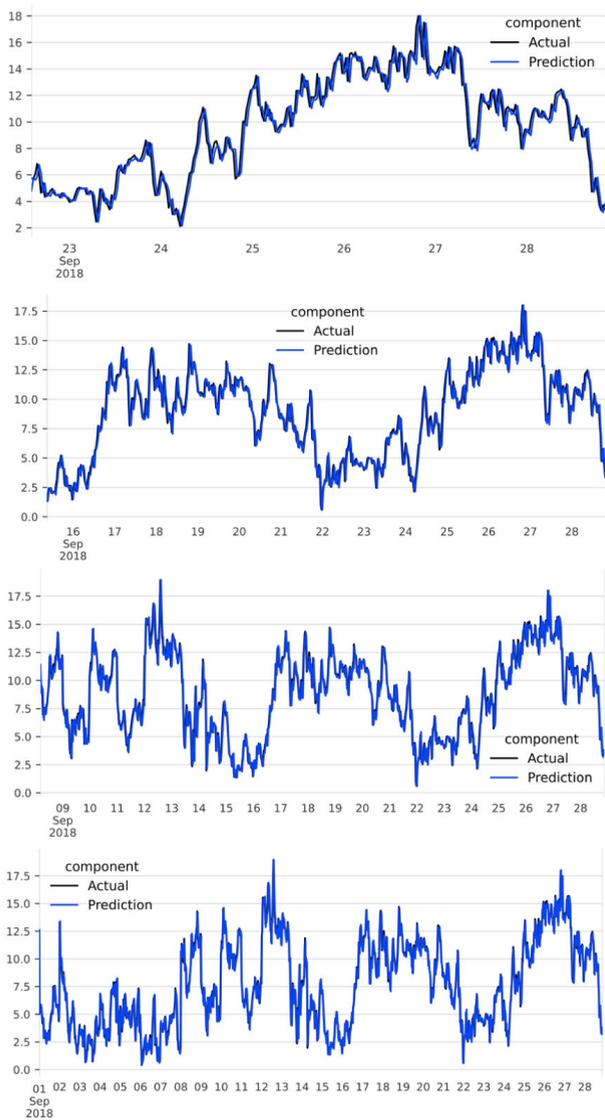


FIGURE 9. Third quarter prediction for 7, 14, 21 and full month.

Fig. 7 presents the results of the TCN model. It shows the predictions are close enough but have a difference with actual wind speed. This difference is measured in terms of performance metrics and reported in Table 3. The second quarter includes the month of April, May, and June, while it predicts the 7, 14, 21, and 30 days as shown in Fig. 8.

TABLE 3. Performance measure for first quarter to predict the wind speed generation.

Metrics	One Week (Q1)	Two Weeks (Q1)	Three Weeks (Q1)	Full Month (Q1)
MAE	1.36	1.71	1.69	1.58
RMSE	1.76	2.17	2.13	2.05

Fig. 8 presents a close prediction to the actual wind speed. The performance measure reported in Table 4 confirms the low error rate of our model prediction. Similarly,

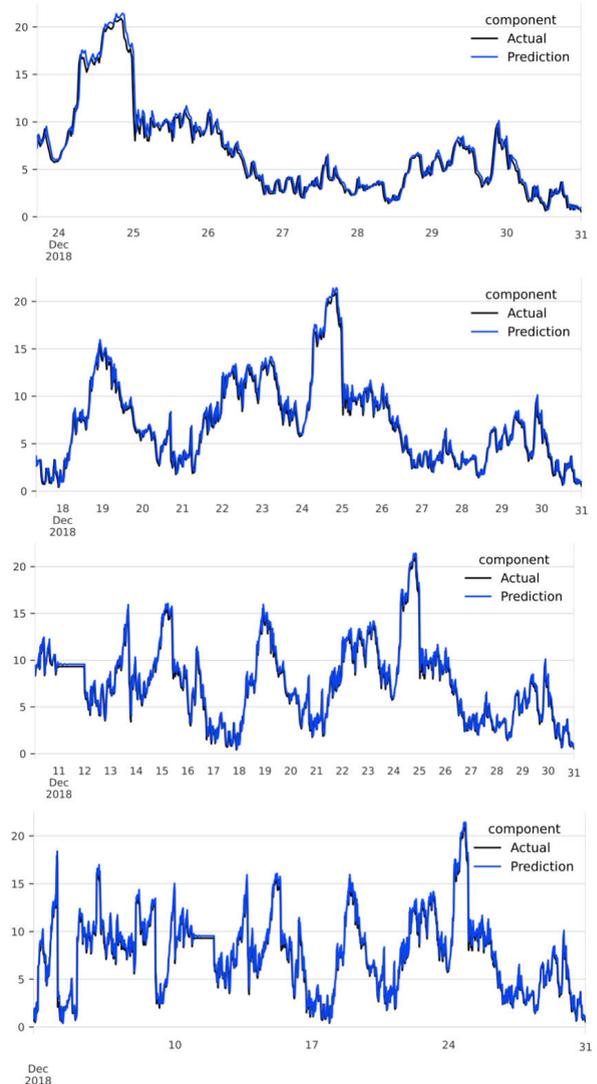


FIGURE 10. Fourth quarter prediction for 7, 14, 21 and full month.

Q3 and Q4 results are presented in Fig. 9, 10 and performance measure in Table 5, 6. The obtained results shows the TCN model learn the wind speed correctly and obtained performance measures results confirms it.

TABLE 4. Performance measure for second quarter to predict the wind speed generation.

Metrics	One Week (Q2)	Two Weeks (Q2)	Three Weeks (Q2)	Full Month (Q2)
MAE	0.88	0.84	0.84	0.86
RMSE	1.25	1.28	1.29	1.40

**B. POWER GENERATION PREDICTION**

The power generation prediction is based on *k*NN regression model. The optimal value of *k* is searched using the grid search as shown in Fig. 12. The x-axis presents the value of *k*

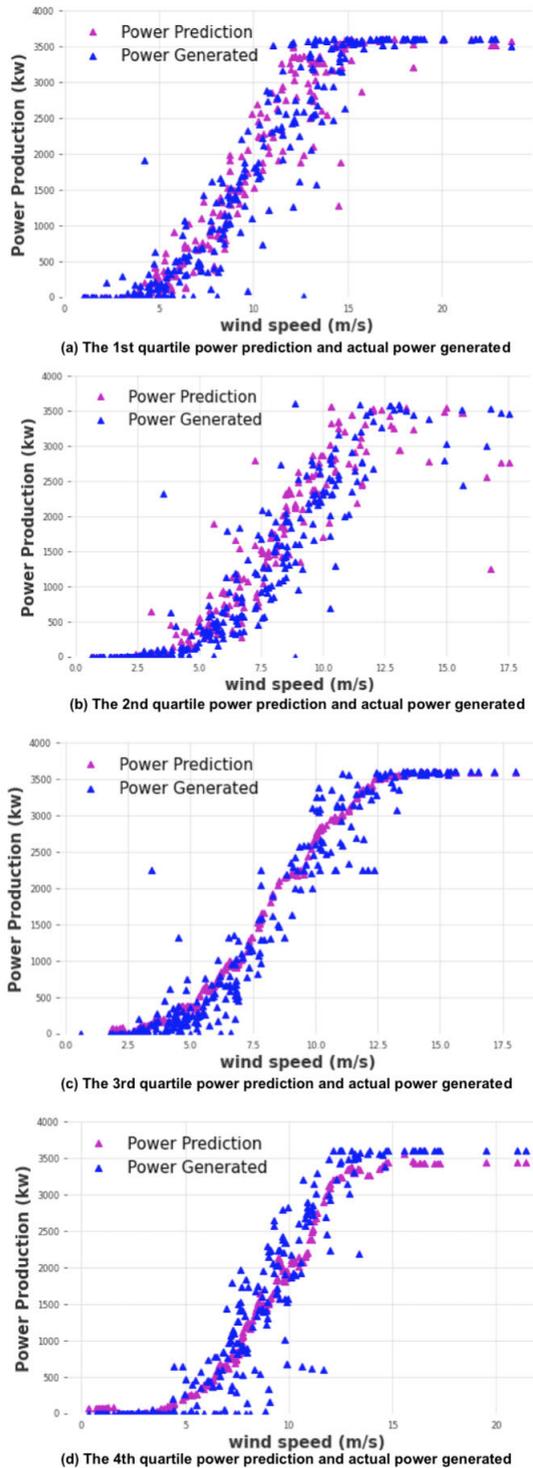


FIGURE 11. The power prediction and actual power generated for each quarter Q1, Q2, Q3, and Q4 from left to right using kNN regression.

that means required neighbors and y-axis presents the mean absolute error.

The forecasting of wind speed over the period of a month is used to predict power generation. The results of kNN regression are presented in Fig. 11.

The obtained results are close enough to the actual power generation. The developed model not only predicts the wind

TABLE 5. Performance measure for third quarter to predict the wind speed generation.

Metrics	One Week (Q3)	Two Weeks (Q3)	Three Weeks (Q3)	Full Month (Q3)
MAE	0.67	0.65	0.69	0.69
RMSE	0.88	0.85	0.93	0.99

TABLE 6. Performance measure for fourth quarter to predict the wind speed generation.

Metrics	One Week (Q4)	Two Weeks (Q4)	Three Weeks (Q4)	Full Month (Q4)
MAE	0.63	0.69	0.68	0.71
RMSE	0.90	0.93	0.93	1.09

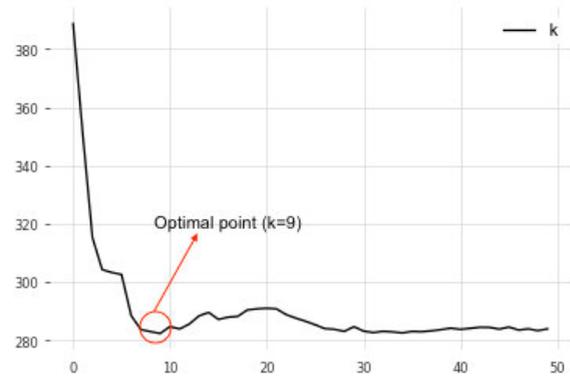


FIGURE 12. The optimal value of k using the grid search.

TABLE 7. Performance measures for power prediction.

Model	Jan - Mar (Q1-MAE)	Apr - Jun (Q2-MAE)	Jul - Aug (Q3-MAE)	Sep - Dec (Q4-MAE)
Our Model	263.59	230.87	226.49	295.46
DT Regression	380.12	349.86	281.64	444.27
Random Forest	326.89	294.23	247.79	389.92
SV Regression	390.72	497.43	307.93	460.45

speed but also predicts power generation. It can help the management team know in advance about the generated power and plan the storage in smart grids. Our objective is to design and develop an advanced pipeline to predict the power generation that can be used for an energy management team to make a timely decision.

### C. COMPARATIVE ANALYSIS

To validate the performance of our model, we compared the results with three state-of-the-art machine learning models. First, the decision tree is a successful approach in predictive modeling as a supervised learning approach. It constructs the tree sequentially by calculating the entropy of the attributes. The theoretical background is rooted in information theory. During the learning phase of the decision tree it splits the nodes by adjusting the numerical parameter of threshold function [29]. Second, the random forest is a

```

1  {
2  "id": "dtmi:ADT:Windfarm;1",
3  "@type": "Interface",
4  "displayName": "Wind Farm",
5  "@context": "dtmi:dtdl:context;2",
6  "contents": [
7    {
8      "name": "contains",
9      "@type": "Relationship",
10     "displayName": "contains",
11     "properties": [
12       {
13         "name": "ownerDataAnalyst",
14         "@type": "Property",
15         "schema": "string"
16       },
17       {
18         "name": "ownershipDepartment",
19         "@type": "Property",
20         "schema": "string"
21       }
22     ]
23   }
24 ]
25 }

1  {
2  "id": "dtmi:windfarm:Windturbine;1",
3  "@type": "Interface",
4  "displayName": "WT - SCADA Unit",
5  "@context": "dtmi:dtdl:context;2",
6  "contents": [
7    {
8      "name": "windSpeed",
9      "@type": "Property",
10     "schema": "double"
11   },
12   {
13     "name": "windDirection",
14     "@type": "Property",
15     "schema": "double"
16   },
17   {
18     "name": "generatedPower",
19     "@type": "Property",
20     "schema": "double"
21   },
22   {
23     "name": "theoreticalPower",
24     "@type": "Property",
25     "schema": "double"
26   }
27 ]
28 }

```

**FIGURE 13.** A DTDL of wind farm and wind turbines to define the relationship and attributes for real-time monitoring.

powerful machine learning algorithm that is based on ensemble learning paradigm. Several trees are constructed over the training dataset known as base learners. A simple voting

mechanism is used to combine the individual results. This can help to reduce the variance of the model which shows high accuracy and robustness in many application domains [30]. Third, support vector regression (SVR) based on statistical theory and successfully applied to finance, forecasting electricity prices, and power consumption [31]. The basic idea of SVR is to transform the input data points to a higher dimensional feature space through a function and separate the feature space with maximum margin. The mean absolute error as a performance measure is reported in Table 7.

Table 7 shows a significant improvement in each quarter which confirms the applicability to predict the power generation.

#### IV. CONCLUSION

Wind farms are contributing towards the generation of clean and affordable energy to support sustainable solutions. The wind-turbines condition monitoring and power generation prediction play an essential role in supporting the management team in making informed decisions and supporting the domestic supply chain. We developed a 5G-NG-RAN assisted cloud-based digital twins framework to monitor the SCADA units of wind turbines. The digital twins enabled real-time monitoring of wind farms without visiting them physically. 5G-NG-RAN assisted cloud allows low latency services to support digital twin in real-time for predictive modeling in wind turbines. Furthermore, a machine learning pipeline is designed over the temporal convolutional neural network and kNN regression to forecast wind and power generation. The empirical evaluation of the publicly available dataset confirms the applicability in real wind-farms scenarios.

Our future plan is to extend this digital twins framework for offshore wind farms monitoring and prediction to support sustainable energy solutions.

#### APPENDIX

Fig. 13 presents the wind farm and wind turbine in digital twins definition language. The properties of SCADA units are defined as contents for real-time monitoring in the virtual wind farm.

#### REFERENCES

- [1] K. G. Logan, J. D. Nelson, B. C. McLellan, and A. Hastings, "Electric and hydrogen rail: Potential contribution to net zero in the U.K.," *Transp. Res. D, Transp. Environ.*, vol. 87, Oct. 2020, Art. no. 102523.
- [2] Y. Wang, Y. Li, R. Zou, A. M. Foley, D. Al Kez, D. Song, Q. Hu, and D. Srinivasan, "Sparse heteroscedastic multiple spline regression models for wind turbine power curve modeling," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 191–201, Jan. 2021.
- [3] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 452–460, Feb. 2014.
- [4] S. Shokrzadeh, M. J. Jozani, and E. Bibeau, "Wind turbine power curve modeling using advanced parametric and nonparametric methods," *IEEE Trans. Sustain. Energy*, vol. 5, no. 4, pp. 1262–1269, Oct. 2014.
- [5] J. C. de Albuquerque, R. R. B. de Aquino, O. N. Neto, M. M. S. Lira, A. A. Ferreira, and M. A. de Carvalho, "Power curve modelling for wind turbine using artificial intelligence tools and pre-established inference criteria," *J. Modern Power Syst. Clean Energy*, vol. 9, no. 3, pp. 526–533, 2021.

- [6] I. Delgado and M. Fahim, "Wind turbine data analysis and LSTM-based prediction in SCADA system," *Energies*, vol. 14, no. 1, p. 125, Dec. 2020.
- [7] J. Tautz-Weinert and S. J. Watson, "Using SCADA data for wind turbine condition monitoring—A review," *IET Renew. Power Gener.*, vol. 11, no. 4, pp. 382–394, 2017.
- [8] Y. Liu, Z. Wu, and X. Wang, "Research on fault diagnosis of wind turbine based on SCADA data," *IEEE Access*, vol. 8, pp. 185557–185569, 2020.
- [9] X. Wu, W. Liu, N. Wang, and Y. Ma, "Short-term wind power prediction based on time series analysis model," in *Proc. 2nd Int. Conf. Comput. Sci. Electron. Eng. (ICCSEE)*, 2013, pp. 1–3.
- [10] H. Sun, C. Qiu, L. Lu, X. Gao, J. Chen, and H. Yang, "Wind turbine power modelling and optimization using artificial neural network with wind field experimental data," *Appl. Energy*, vol. 280, Dec. 2020, Art. no. 115880.
- [11] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," *Appl. Energy*, vol. 224, pp. 550–566, Aug. 2018.
- [12] Y. Hu, Y. Qiao, J. Liu, and H. Zhu, "Adaptive confidence boundary modeling of wind turbine power curve using SCADA data and its application," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1330–1341, Jul. 2019.
- [13] Y. Wu, K. Zhang, and Y. Zhang, "Digital twin networks: A survey," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13789–13804, Sep. 2021.
- [14] L. Lei, G. Shen, L. Zhang, and Z. Li, "Toward intelligent cooperation of UAV swarms: When machine learning meets digital twin," *IEEE Netw.*, vol. 35, no. 1, pp. 386–392, Jan. 2021.
- [15] J. Lopez, J. E. Rubio, and C. Alcaraz, "Digital twins for intelligent authorization in the B5G-enabled smart grid," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 48–55, Apr. 2021.
- [16] W. Sun, H. Zhang, R. Wang, and Y. Zhang, "Reducing offloading latency for digital twin edge networks in 6G," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12240–12251, Oct. 2020.
- [17] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Communication-efficient federated learning for digital twin edge networks in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5709–5718, Aug. 2021.
- [18] A. Ghasempour, "Internet of Things in smart grid: Architecture, applications, services, key technologies, and challenges," *Inventions*, vol. 4, no. 1, p. 22, Mar. 2019.
- [19] O. O. Olatunji, P. A. Adedeji, N. Madushele, and T.-C. Jen, "Overview of digital twin technology in wind turbine fault diagnosis and condition monitoring," in *Proc. IEEE 12th Int. Conf. Mech. Intell. Manuf. Technol. (ICMIMT)*, May 2021, pp. 201–207.
- [20] K. Sivalingam, M. Sepulveda, M. Spring, and P. Davies, "A review and methodology development for remaining useful life prediction of offshore fixed and floating wind turbine power converter with digital twin technology perspective," in *Proc. 2nd Int. Conf. Green Energy Appl. (ICGEA)*, Mar. 2018, pp. 197–204.
- [21] J. Kim, D. G. Duguma, P. V. Astillo, H.-Y. Park, B. Kim, I. You, and V. Sharma, "A formally verified security scheme for Inter-gNB-DU handover in 5G vehicle-to-everything," *IEEE Access*, vol. 9, pp. 119100–119117, 2021.
- [22] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, "Digital twin driven prognostics and health management for complex equipment," *CIRP Ann.*, vol. 67, no. 1, pp. 169–172, 2018.
- [23] F. Tao, Q. Qi, L. Wang, and A. Nee, "Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: Correlation and comparison," *Eng. Forthcoming*, vol. 5, no. 4, pp. 653–661, 2019.
- [24] P. Lara-Benitez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, "Temporal convolutional networks applied to energy-related time series forecasting," *Appl. Sci.*, vol. 10, no. 7, p. 2322, Mar. 2020.
- [25] R. Goyal, P. Chandra, and Y. Singh, "Suitability of KNN regression in the development of interaction based software fault prediction models," *IERI Proc.*, vol. 6, pp. 15–21, Jan. 2014.
- [26] B. Erisen. (2018). *Wind Turbine SCADA Dataset*. [Online]. Available: <https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset>
- [27] D. Astolfi, F. Castellani, A. Lombardi, and L. Terzi, "Multivariate SCADA data analysis methods for real-world wind turbine power curve monitoring," *Energies*, vol. 14, no. 4, p. 1105, Feb. 2021.
- [28] D. Karamichailidou, V. Kaloutsas, and A. Alexandridis, "Wind turbine power curve modeling using radial basis function neural networks and Tabu search," *Renew. Energy*, vol. 163, pp. 2137–2152, Jan. 2021.
- [29] S. Mitrofanov and E. Semenkina, "An approach to training decision trees with the relearning of nodes," in *Proc. Int. Conf. Inf. Technol. (InfoTech)*, Sep. 2021, pp. 1–5.
- [30] Z. Chai and C. Zhao, "Multiclass oblique random forests with dual-incremental learning capacity," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5192–5203, Dec. 2020.
- [31] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.



**MUHAMMAD FAHIM** received the B.S. degree (Hons.) from Gomal University, Pakistan, in 2007, the M.S. degree from the National University of Computer and Emerging Sciences (NUCES), Pakistan, in 2009, and the Ph.D. degree from Kyung Hee University, South Korea, in February 2014. He worked as a Postdoctoral Fellow at the Department of Computer Engineering, Kyung Hee University. He was an Assistant Professor with the Department of Computer and Software Engineer-

ing, Faculty of Engineering and Natural Sciences, Istanbul Sabahattin Zaim University, Istanbul, Turkey, for three years. He also led the Machine Learning Research Laboratory, Istanbul Sabahattin Zaim University. He worked as an Assistant Professor at the Institute of Information Systems, Innopolis University, Innopolis, Russia, for four years. He is currently a Lecturer with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, U.K. His research interests include machine learning models, wearable computing, digital signal processing, and behavior recognition in intelligent environments.



**VISHAL SHARMA** (Senior Member, IEEE) received the B.Tech. degree in computer science and engineering from Punjab Technical University, in 2012, and the Ph.D. degree in computer science and engineering from Thapar University, India, in 2016. From November 2016 to March 2019, he was with the Information Security Engineering Department, Soonchunhyang University, South Korea, in multiple positions (from November 2016 to December 2017, he was a Postdoctoral Researcher, and from January 2018 to March 2019, he was a Research Assistant Professor). He was also a joint Postdoctoral Researcher with Soongsil University, South Korea. Before this, he worked as a Lecturer with the Department of Computer Science and Engineering, Thapar University. He is currently working as a Lecturer (Assistant Professor) with the School of Electronics, Electrical Engineering and Computer Science (EECS), Queen's University Belfast (QUB), Belfast, U.K. At QUB, he is also a Racial Equity Network School Co-Champion, a BCS Accreditation Liaison, and a Mental Health Ambassador (MHFA). Before coming to QUB, he was a Research Fellow at the Information Systems Technology and Design (ISTD) Pillar, Singapore University of Technology and Design (SUTD), Singapore, where he worked on the future-proof blockchain systems. He has authored/coauthored more than 100 journals/conference articles and book chapters and co-edited two books with Springer. His research interests include autonomous systems, UAV communications, network behavior modeling, 5G and beyond, blockchain, and CPS security. He is a Professional Member of ACM and the Past Chair of ACM Student Chapter-TIET Patiala. He was a recipient of four best paper awards. He serves on the Editorial Board of *IEEE Communications Magazine*, *CAAI Transactions on Intelligence Technology (IET)*, *Wireless Communications and Mobile Computing*, *IET Networks*, and *ICT Express*, and the Section Editor-in-Chief of *Drones* journal. He is also the Interim Co-Chair of the IEEE U.K. and Ireland Diversity, Equity, and Inclusion Committee.



**TUAN-VU CAO** received the Ph.D. degree in electrical engineering from the University of Oslo, in 2012. He has a long experience in both academic and industry. From 2012 to 2018, he worked as a Postdoctoral Fellow at the Norwegian University of Science and Technology and a Senior Engineer at WINS Instrumentation AS and Prediktor AS for various projects, such as monitoring nutrient food content, wireless DST real time instrumentation, and micro-power sensor interface in nanometer

CMOS technology. Since 2018, he joined NILU. He is currently a Senior Scientist at the Norwegian Institute for Air Research (NILU). He is a Co-PI of EEA Grant Project—HAPADS (2020–2023)—a novel air mobile monitoring system enables end-users to make information-driven decisions to mitigate air pollution exposure. Key Designer and Technical Coordinator of NFR 273394-“Leopard-Wearable particle detector enabling safer working environments” (2018–2020). His research interests include enabling technologies and autonomous systems for environmental monitoring and management.



**BERK CANBERK** (Senior Member, IEEE) is currently a Professor with the Department Artificial Intelligence and Data Engineering, Istanbul Technical University. Since 2016, he has been an Adjunct Professor with the Department of Electrical and Computer Engineering, Northeastern University. His current research interests include AI driven network automation and management, software-defined networking, 5G, 6G, and intelligent aerial networks. He was a recipient of the

IEEE Turkey Research Incentive Award, in 2018, the IEEE INFOCOM Best Paper Award, in 2018, the British Council (United Kingdom) Researcher Link Award, in 2017, the IEEE CAMAD Best Paper Award, in 2016, and the IEEE INFOCOM Best Poster Paper Award, in 2015.



**TRUNG Q. DUONG** (Fellow, IEEE) is currently the Chair Professor of telecommunications at Queen’s University Belfast, Belfast, U.K. and a Research Chair of the Royal Academy of Engineering. He has authored or coauthored over 400 publications. His current research interests include wireless communications, machine learning, real-time optimization, and data analytic. He serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and

IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and an Executive Editor for IEEE COMMUNICATIONS LETTERS. He has served as an Editor/Guest Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, *IEEE Communications Magazines*, IEEE COMMUNICATIONS LETTERS, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was awarded the Best Paper Award from the IEEE Vehicular Technology Conference (VTC-Spring), in 2013, the IEEE International Conference on Communications (ICC), in 2014, the IEEE Global Communications Conference (GLOBECOM), in 2016 and 2019, the IEEE Digital Signal Processing Conference (DSP), in 2017, and the International Wireless Communications and Mobile Computing Conference (IWCMC), in 2019. He was also a recipient of the prestigious Royal Academy of Engineering Research Fellowship, from 2015 to 2020) and a prestigious Newton Prize, in 2017.

...