A Data-Mining Approach to Monitoring Wind Turbines

Andrew Kusiak, *Member, IEEE*, and Anoop Verma, *Student Member, IEEE*

***Abstract*—The rapid expansion of wind farms has generated interest in operations and maintenance. An operating wind turbine undergoes various state changes, including transformation from a normal to a fault mode. Condition-based maintenance tools are needed to identify potential faults in the system. The prediction of turbine fault modes is of particular interest. In this research, data-mining algorithms are employed to construct prediction models for wind turbine faults. A three-stage prediction process is followed: 1) prediction of a fault of any kind; 2) prediction of specific faults of the system; and 3) identification on unseen faults. A comparative analysis of various data-mining algorithms is reported based on the data collected at a large wind farm. Random forest algorithm models provided the best accuracy among all algorithms tested. The robustness of the predictive model is validated for faults that have occurred at turbines with previously unseen data. The research results discussed in this paper have been derived from data collected at 17 wind turbines.**

***Index Terms*—Data mining, multiclass classification, prediction, wind turbine, wind turbine states.**

I. Introduction

WIND energy is regarded as a major renewable resource destined to grow in importance in the decades to come. The expansion of wind farms makes their operations and maintenance (O&M) an important issue. It is not unusual for the maintenance/repair cost of wind turbine components to exceed their procurement cost [1], [2]. According to the data presented in [3], maintenance cost alone may account for at least 10% of the total generation cost. To address O&M issues, traditional maintenance practices such as periodic and corrective maintenance are being replaced with condition-based monitoring and maintenance.

State-of-the-art condition maintenance applications in the wind industry are discussed in [4]–[7]. Condition-based monitoring approaches continuously monitor the performance of wind turbine components with installed sensors and equipment. Vibration analysis [8], optical strain measurements [9], and oil particle analysis [10] are commonly used in condition monitoring. Performance monitoring is another promising approach that closely resembles condition monitoring. It utilizes historical wind turbine data to predict wind turbine performance parameters such as gearbox oil temperature, and tower acceleration. Performance monitoring is a cost-effective approach to analyze wind turbine performance as the Supervisory Control and Data Acquisition (SCADA) system records various wind turbine parameters that could be fault informative.

Data mining has been used as a viable approach to performance monitoring of wind turbines. Related data-mining algorithm applications include fault diagnosis [1], [11], modeling of abnormal behavior [12], [13], and power curve monitoring [14]. Other related research includes identification of status patterns of wind turbines [15], in which the authors employed association rule mining to identify patterns within the individual statuses. Here the term status represents a potential fault. In reference [16], the authors employed an adaptive control strategy to gain maximum power and minimum torque ramp.

Considering the role of converters in optimizing wind turbine performance, a stream of research has focused on reliability assessment of wind turbines [17], [18].

The research reported in this paper utilized data-mining algorithms to predict wind turbine states. The results presented in the paper are based on the analysis of data obtained from 17 wind turbines (1.5 MW) of a large wind farm in Blairsburg, Iowa. The values of parameters recorded at 10-s intervals (10-s data) over a four-month period constitute the dataset for this research.

This paper is organized as follows. Section II describes wind turbine states along with a discussion of the data preprocessing. Section III provides the computational results. Section IV concludes the research and presents future research directions.

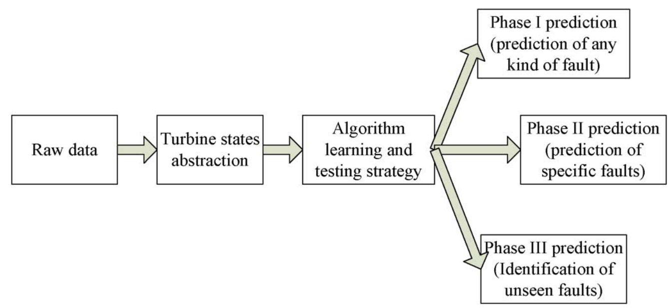

Fig. 1. Framework of the proposed approach.

II. Models for Monitoring Wind Turbine States

The framework for building prediction models is provided in Fig. 1. An abstraction of turbine states is used to categorize the output data into a number of states using expert knowledge. Model building involves using various data-mining algorithms. The models are then tested. The generated dataset is used to construct models for Phase-I and Phase-II predictions. The main

Manuscript received March 21, 2011; revised June 08, 2011; accepted July 21, 2011. Date of publication July 29, 2011; date of current version December 16, 2011. This work was supported by funding from the Iowa Energy Center under Grant 07-01.

The authors are with the Intelligent Systems Laboratory, The University of Iowa, Iowa City, IA 52242-1527 USA (e-mail: andrew-kusiak@uiowa.edu; anoop-verma@uiowa.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2011.2163177

TABLE I
TURBINE STATE INFORMATION

State Number	State Description
1	Turbine OK with no errors
2	Turbine running smoothly
3	Turbine running up idling for cut in
4	Turbine in maintenance mode
5	Turbine in repair mode
6	Power failure/grid downtime
7	Weather downtime
8	Turbine stopped externally
9	Turbine stopped locally
10	Turbine stopped remotely
11	Emergency stop
12	Turbine stopped due to curtailment
13	Turbine stopped by customer
14	Turbine idling locally
15	Turbine idling remotely
16	Wind direction curtailment
17*	Turbine in fault mode
*Actual fault	

objective of Phase-I is to predict a fault of any kind, whereas, predictions in Phase-II target specific faults. In Phase-III predictions, unseen faults from different wind turbines are identified. Descriptions of various wind turbine states are provided next.

A. Turbine State Description

The variability of wind speed impacts the performance of wind turbines and is recorded as fault states. Normal operations, weather-related downtime, maintenance downtime, fault mode, and emergency stop are some of the many states recorded by the SCADA system of a wind turbine.

States changes may vary from insignificant (e.g., when a turbine is changing its state from idle to normal operations) to a potential fault. Table I lists the 17 possible states of a wind turbine. State number 17 represents the fault mode of wind turbines and there can be more than 400 possible ways in which a wind turbine can be faulted. Gearbox oil over-temperature, blade angle asymmetry, pitch thyristor fault, and yaw runaway are some of the common fault modes of a wind turbine. In the research reported in this paper, the main emphasis was to predict the fault mode of wind turbine ahead of actual occurrence.

B. Abstraction of Turbine States

A typical turbine may undergo a number of different states including turbine normal operations, run-up idling, maintenance/repair mode, fault mode, weather downtime, etc. The prediction of a turbine's fault mode is of particular interest as it represents some potential fault in the system. A turbine in state 17 can be affected by as many as 400 different fault modes of varying intensity. Fig. 2 shows the histogram of 17 wind turbines plotted over a period of four months (from 8/27/2010 to 12/4/2010). Based on the frequency of fault mode, turbine 12 was considered in the analysis. In order to reduce the computational effort required by data-mining algorithms, the recorded states of wind turbines were further categorized using domain knowledge. Table II represents the initially recorded and categorized states of turbine 12. The initial 44 turbine states were categorized into four states: *Turbine OK*, *Fault*, *Weather downtime*, and *Maintenance downtime*. The *Turbine OK* category

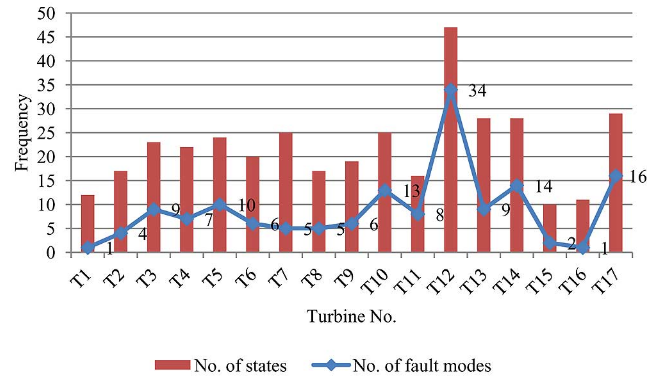

Fig. 2. Comparison of wind turbines states.

corresponds to normal functioning operations, including run-up idling, whereas, the *Fault* category corresponds to an actual or potential fault in the system. The *Weather downtime* category corresponds to turbine downtime due to poor weather conditions, whereas, any other downtime is considered as *Maintenance downtime*.

C. Learning Strategy

For both prediction phases, the dataset was divided into two parts, i.e., initial dataset and blind dataset. The data-mining algorithms used two-thirds of the initial data for training, and the remaining one-third of the initial data was used for testing. The performance of the data-mining algorithms on the test dataset was used for algorithm selection. The best performing algorithm was then used to construct prediction models on the blind dataset (discussed in Section III). Details regarding the parameter selection are discussed in the next section.

1) Parameter Selection: A Supervisory Control and Data Acquisition (SCADA) system records more than 100 wind turbine parameters that can be broadly categorized into: 1) wind turbine performance parameters, 2) wind turbine control parameters, and 3) wind turbine noncontrollable parameters. Parameters such as power, generator speed, and rotor speed are the performance parameters, whereas, blade pitch angle and generator torque are controllable parameters. Wind speed is the only noncontrollable parameter. In the research reported in this paper, a combination of turbine performance parameters, control parameters, and noncontrollable parameters are used to predict the wind turbine states. To minimize the data dimensionality and to remove irrelevant parameters, parameter selection algorithms are used. A month of data was used for parameter selection and algorithm learning. A stratified subset of the original data was used for parameter selection to make the process computationally efficient. Fig. 3 displays the original and stratified data. Distribution of the output class is preserved in stratified data to avoid bias towards any specific class. Three different data-mining algorithms, wrapper with genetic search (WGS) [19], [20], wrapper with best first search (WBFS) [21], and boosting tree algorithm (BTA) [22] were selected to determine relevant parameters for prediction of turbine states. Wrapper is a supervised learning approach using different search techniques to select the relevant parameters by performing ten-fold cross validation. Table III lists the ten best parameters from each pa-

TABLE II
TURBINE STATE CATEGORIES

Fault Mode States		States Other than Fault Mode	
States description	Output class	States description	Output class
Asymmetric generator current	Fault	Online	Turbine OK
Axle 1 fault pitch controller	Fault	Run-up idling	Turbine OK
Battery charging rotor blade drive	Turbine OK	Maintenance mode	Maintenance downtime
Cable twisting left	Turbine OK	Repair mode	Maintenance downtime
Cable twisting right	Turbine OK	Grid downtime	Maintenance downtime
Centrifugal switch	Fault	Turbine curtailment	Turbine OK
Gearbox oil over temperature	Fault	Stopped externally	Maintenance downtime
Gearbox oil temperature too low	Weather downtime	Stopped locally	Maintenance downtime
Hydraulic pump time too high	Fault	Stopped remotely	Maintenance downtime
Limit switch 90°-rotor blade defective	Maintenance downtime	Weather conditions	Weather downtime
Maintenance switch pitch	Maintenance downtime		
Maximum motor power	Fault		
Pitch overrun 0°	Fault		
Pitch thyristor fault	Fault		
Pulse sensor rotor monitor defect	Fault		
Reply generator high stage	Fault		
Temperature warning pitch motor	Turbine OK		
Wrong parameter check sum	Turbine OK		

TABLE III
SELECTED PARAMETERS USING DATA-MINING ALGORITHMS

No.	WGS	WBFS	BTA
	10 fold cross validation (%)	10 fold cross validation (%)	Parameter importance (%)
1	Nacelle revolution (90)*	Blade 3 pitch angle (actual) (100)*	Blade2 pitch angle (actual) (100)*
2	Blade 3 pitch angle (actual)(90)*	Current phase C (80)*	Blade3 pitch angle (actual) (95)*
3	Current Phase B (70)	Temperature hub (80)*	Blade1 pitch angle (actual) (94)*
4	Nacelle Position (70)	Temp. control box axis 1 (60)	Generator/gearbox speed (86)*
5	Generator/gearbox speed (70)*	Voltage phase C (50)	Generator speed (85)
6	Temperature, bearing B (70)	Generator speed (50)	Rotor speed (70)
7	Temperature top box (°C) (70)	Drive train acceleration (50)	Blade2 pitch angle (set) (68)
8	Power (Actual) (60)	Temperature top box (50)	Blade3 pitch angle (set) (68)
9	Tower deflection (60)	Nacelle revolution (40)*	Blade1pitch angle (set) (68)
10	Wind deviation, 1 sec (60)	Temperature bearing A (40)	Drive train acceleration (65)
*Selected parameters			

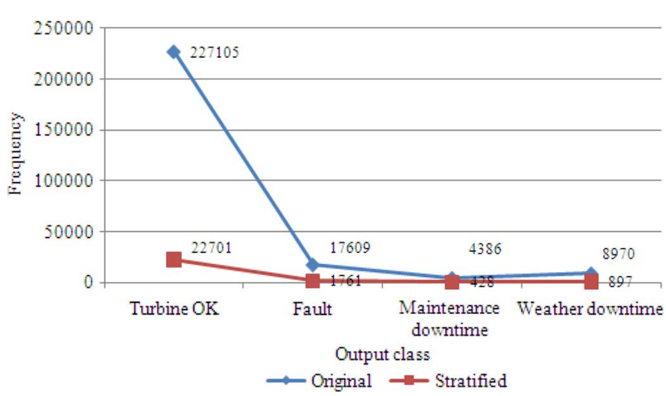

Fig. 3. Output class distribution.

rameter selection algorithm. Parameters for nacelle revolution, blade (1–3) pitch angle, current Phase C, temperature hub, and generator/gearbox speed were finally selected to build the prediction models.

2) Evaluation Metric: The evaluation of data-mining algorithms is based on the prediction accuracy of each output class. Considering the imbalance in an output class, geometric mean (gmean) of the output class is used as criteria for selecting data-mining algorithms for the prediction task. The evaluation of accuracy is presented in a confusion matrix (see Fig. 4).

Equation (1) defines the geometric mean (gmean) of the output class, whereas, acc_i is the accuracy of class i, n is the total number of output classes

$$\mathrm{gmean} = \sqrt{\prod_{i=1}^{n} \mathrm{acc}_i}. \quad (1)$$

3) Algorithm Selection: Five data-mining algorithms: neural network (NN), support vector machine (SVM), random forest algorithm (RFA), boosting tree algorithm (BTA), and general chi-square automatic interaction detector (CHAID) algorithm were initially selected for building models at t time stamp. NN uses backpropagation to classify instances [23]. Twenty NN models with different kernels and structures were built in this research, and the most accurate and robust model was selected. SVM constructs a hyperplane or set of hyperplanes in a high dimensional space, which can be used for classification, or regression. In SVM the hyperplane with the largest distance to the nearest training data points of any class (so-called functional margin), yields good accuracy [24]. RFA is an ensemble

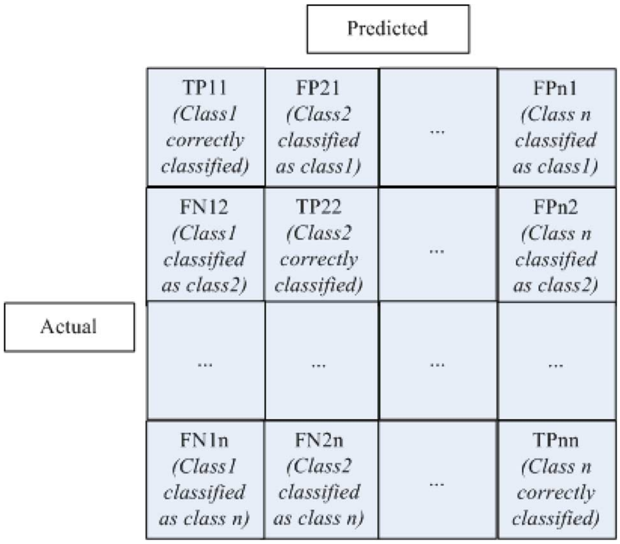

Fig. 4. Confusion matrix.

TABLE IV
PHASE-I PREDICTION RESULTS

Algorithm	Output Class: Turbine OK [%]	Fault [%]	Maintenance downtime [%]	Weather downtime [%]	Overall Accuracy [%]
SVM	99.1	91.0	30.4	54.5	95.8
CHAID	99.0	89.6	35.5	67.1	96.0
NN	99.5	93.5	62.3	84.9	97.6
BTA	99.6	93.2	83.3	96.9	98.8
RFA	99.8	99.6	78.4	97.9	99.4

learning method where multiple random trees are generated during classification. It selects k random input parameters for each node split [25]. BTA generates multiple models and applies a weighted combination of the predictions from individual models to derive a single prediction model [22]. CHAID is a tree-based data-mining algorithm that performs multilevel splits for classification [26].

The prediction accuracy for each class (Phase-I predictions) is provided in Table IV. Essentially, all the algorithms performed well while predicting *Turbine OK* and *Fault* class, however, the output class *Weather downtime* and *Maintenance downtime* were predicted with relatively low accuracy. The geometric mean metric gmean indicates that when all classes are predicted with perfect accuracy its value is 1. The algorithm with the highest value of gmean was selected to build prediction models at different time stamps. From the graph in Fig. 5 and Phase-I prediction results (Table IV), both the boosting tree algorithm and the random forest algorithms outperformed the remaining three data-mining algorithms. However, RFA was selected to build the prediction models, as it possesses great generalization ability and it is almost insensitive to the size of the dataset. Fig. 6 illustrates the tree complexity of the random forest algorithm as a function of the misclassification rate. The optimal number of trees was found to be 91.

The same five algorithms were considered for constructing Phase-II prediction models. The output class *Fault* from the Phase-I prediction was replaced by actual fault type, resulting in

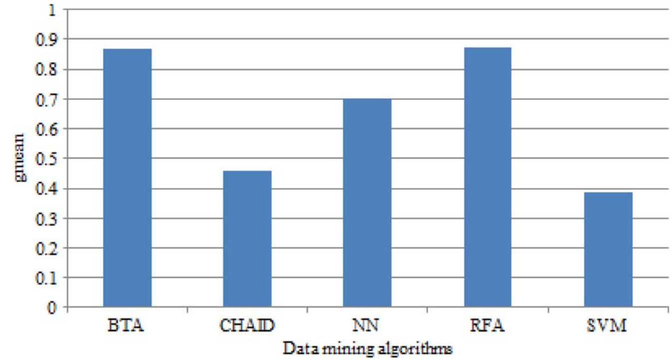

Fig. 5. Performance of different data-mining algorithms using gmean as a criterion.

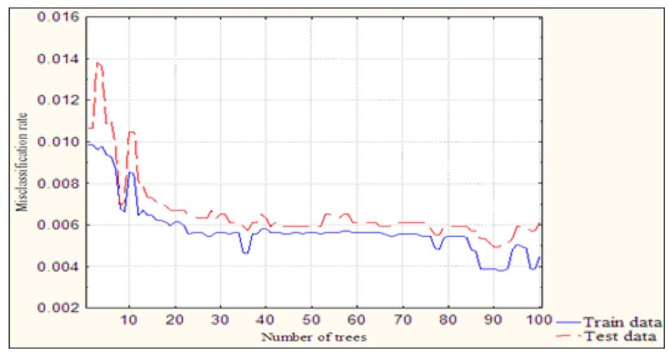

Fig. 6. Misclassification rate of RFA as a function of tree size.

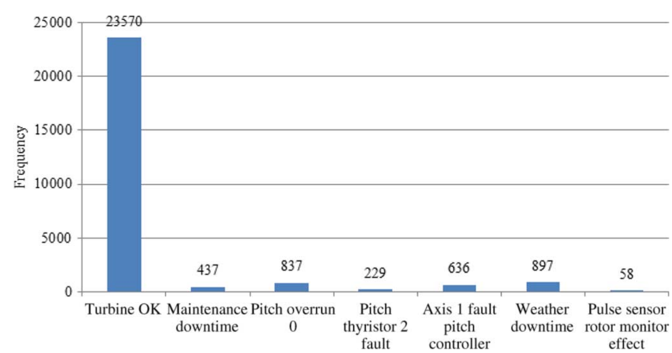

Fig. 7. Distribution of output class at time stamp t.

overall seven output classes. Fig. 7 displays the distribution of data at time stamp t. In the figure, pitch overrun 0° is triggered when limit switch experience a nonpositive angle at least one of the rotor blades. Pitch thyristor 2 fault is triggered when the thyristor is not ready even though the grid conductor is switched ON. Pitch thyristor fault indicate defective axle cabinet. Axle 1 fault pitch controller reports axle disturbance. Pulse sensor rotor monitor defect is due to no pulses to overspeed monitor when the generator over speeds. Table V illustrates the performance of different data-mining algorithms on a t time stamped dataset. It can be seen from Table V that most of the algorithms failed to predict minority output classes (a class with few instances) and thereby resulted in a gmean equal to 0. Only NN and RFA yielded a gmean value greater than 0 (Fig. 8). As anticipated, RFA outperforms the other data-mining algorithms, providing better accuracy for each output class.

TABLE V
PHASE-II PREDICTION RESULTS AT TIME STAMP t

Algorithm	Output Class							Overall Accuracy [%]
	Turbine OK [%]	Maintenance downtime [%]	Weather downtime [%]	Axle 1 fault pitch controller [%]	Pitch overrun 0 [%]	Pitch thyristor 2 fault [%]	Pulse sensor rotor monitor defect [%]	
SVM	99.16	26.82	56.73	51.14	67.12	0.00	0.00	93.08
CHAID	98.98	0.00	97.17	100.0	08.30	30.55	0.00	93.42
NN	99.73	87.19	93.10	99.54	97.23	54.16	37.50	98.65
BTA	99.03	84.96	24.46	35.43	67.37	0.00	0.00	92.88
RFA	**99.64**	**82.70**	**99.29**	**100.0**	**98.93**	**87.83**	**61.90**	**98.83**

TABLE VI
PREDICTION ACCURACY OF OUTPUT CLASS USING RFA (PHASE-I PREDICTION)

Time stamp [s]	Output Class				Overall Accuracy [%]
	Turbine OK [%]	Fault [%]	Maintenance downtime [%]	Weather downtime [%]	
t	99.88	99.67	78.41	97.91	99.45
$t+10$	99.56	99.00	77.22	95.04	98.39
$t+30$	97.64	96.41	74.59	94.62	96.54
$t+60$	95.70	95.64	71.67	92.55	94.43
$t+120$	91.87	90.00	67.49	88.47	90.89
$t+180$	88.58	87.34	64.94	84.43	86.82
$t+240$	85.62	84.64	60.31	82.44	83.93
$t+300$	83.05	82.76	59.67	80.39	81.76

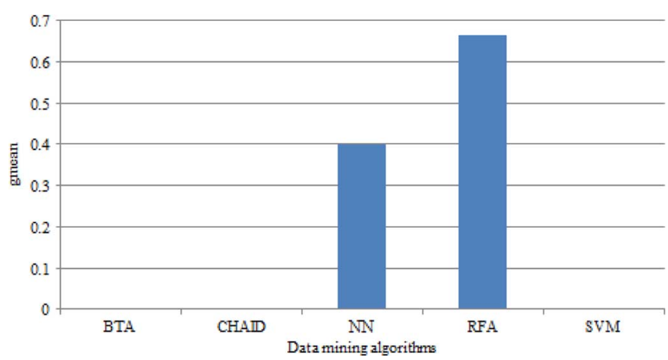

Fig. 8. Performance of different data-mining algorithms using gmean as a criterion (Phase-II prediction).

III. COMPUTATIONAL RESULTS

In this section, the random forest algorithm (RFA) was used to build eight prediction models at various time stamps, with a maximum prediction length of 5 min. The maximum tree size for the random forest algorithm was set to 300. The accuracy was found to be in the range of 81%–99% for all output classes (Table VI).

A. Phase-II Prediction

In this phase, output class *Fault* was replaced with the actual fault types, these being pitch overrun 0°, pitch thyristor 2 fault, axle 1 fault pitch controller, and pulse sensor motor defect.

Table VII displays the prediction results produced by the RFA at different time stamps. The accuracy of each output class was found to be in the range 68%–100%, except for output class pulse sensor rotor monitor defect for which accuracy was low (e.g., 40.67%–61.9%). Fig. 9 presents the gmean value of both prediction phases. Phase-I predictions had overall better gmean values

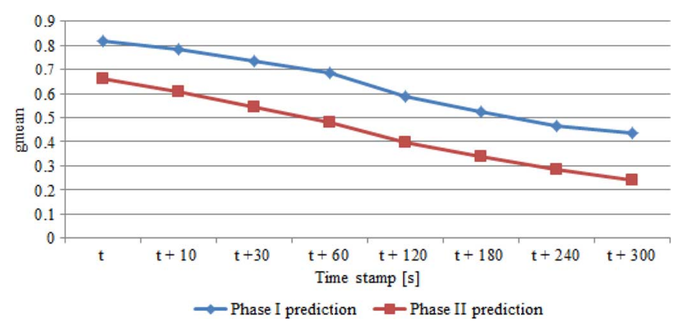

Fig. 9. Values of gmean at various time stamps.

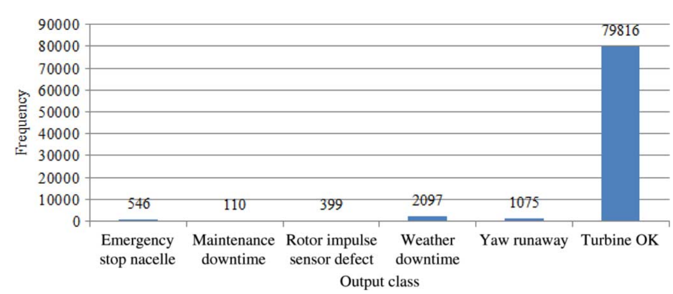

Fig. 10. Distribution of output classes (Turbine 10).

(0.435–0.817) than Phase-II predictions (0.242–0.659), because of the poor accuracy of one output class, the pulse sensor rotor monitor defect. In the next section, Phase-III prediction results are illustrated and unobserved faults are identified.

B. Phase-III Predictions

While the results on the testing dataset indicated the effectiveness of the random forest algorithm, in order to validate the

TABLE VII
PREDICTION ACCURACY OF OUTPUT CLASS USING RFA (PHASE-II PREDICTION)

Time stamp[s]	Output Class							Overall accuracy [%]
	Turbine OK [%]	Maintenance downtime [%]	Weather downtime [%]	Axle 1 fault pitch controller [%]	Pitch overrun 0° [%]	Pitch thyristor 2 fault [%]	Pulse sensor rotor monitor defect [%]	
t	99.64	82.70	99.29	100.0	98.93	87.83	61.90	98.83
t + 10	99.34	81.08	97.84	98.67	97.77	85.43	58.94	96.44
t +30	97.15	79.26	95.23	97.13	95.52	83.29	55.03	94.09
t + 60	95.28	76.87	92.86	95.45	93.26	80.09	51.98	91.68
t + 120	90.10	71.58	88.41	91.90	90.23	77.67	48.83	87.82
t + 180	87.98	68.71	85.73	86.39	86.88	74.87	46.29	84.69
t + 240	84.45	65.32	83.66	82.55	83.34	71.45	43.91	81.53
t + 300	82.76	62.43	81.45	80.76	79.55	68.32	40.67	78.35

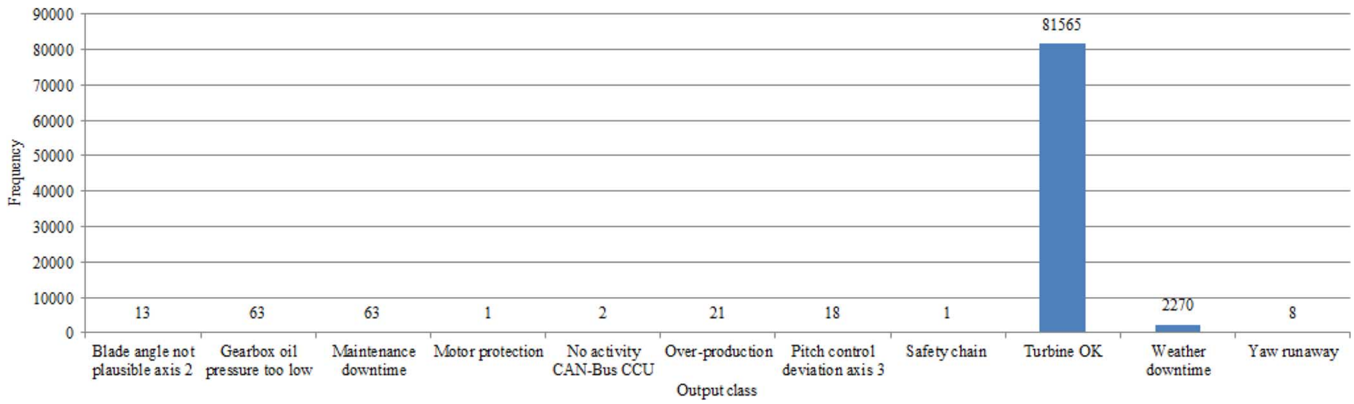

Fig. 11. Distribution of output classes (Turbine 14).

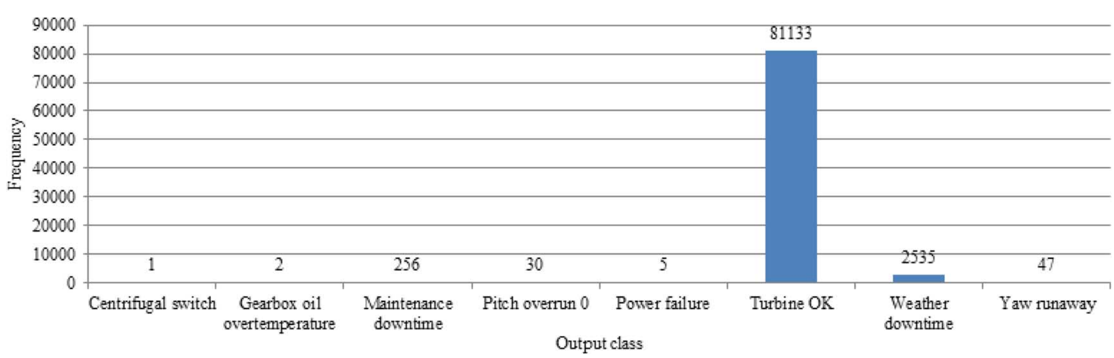

Fig. 12. Distribution of output classes (Turbine 17).

robustness of the proposed model, data from other fault-prone turbines were analyzed with the additional objective of seeing how the model would respond to unseen states types. Due to the inherent variability in wind turbines, faults in a wind turbine vary from one to another. It is interesting to observe the models' responsiveness when some unseen faults are presented. In this section, data from three other fault-prone wind turbines, Turbine 10, Turbine 14, and Turbine 17 are analyzed. Month-long data, from 8/28/2010 until 9/28/2010, were used for the analysis. The models built for Phase-I predictions were deployed for analysis of this dataset. Faults such as yaw runaway, brush wear warning, blade angle implausibility, and reply generator high stage were studied. Figs. 10–12 display the actual distribution of output classes for Turbines 10, 14, and 17, respectively. The numbers of faults vary across turbines, however, the turbines were found to be operating normally with no errors most of the time. Tables VIII–X display the accuracy of output classes across turbines 10, 14, and 17, respectively. It is clear from the results that the algorithms are robust enough to identify unseen faults such as yaw runaway, blade angle not plausible axis 2, etc. The accuracy for correctly identifying unseen fault cases was found to be in the range of 60%–100%, except for faults related to gearboxes (e.g., gearbox over-temperature, gearbox oil pressure too low) which were always identified as *Turbine OK*. The reasons for this include a lack of related input parameters (e.g., gearbox temperature, gearbox oil pressure, etc.) in the model. The results shown in Tables VIII–X confirm that the proposed model can be used to predict most wind turbine faults.

TABLE VIII
MODEL ANALYSIS ON TURBINE 10

Actual output	Anticipated output	Correctly identified cases
Emergency stop nacelle	Fault	85.66%
Maintenance downtime	Maintenance downtime	100%
Rotor impulse sensor defect	Fault	60.90%
Weather downtime	Weather downtime	69.84%
Yaw runaway	Fault	99.62%
Turbine OK	Turbine OK	99.96%

TABLE IX
MODEL ANALYSIS ON TURBINE 14

Actual output	Anticipated output	Correctly identified cases
Blade angle not plausible axis 2	Fault	76.92%
Gearbox oil pressure too low	Fault	0.00%
Maintenance downtime	Maintenance downtime	100%
Motor protection	Fault	100%
No activity CAN-Bus CCU	Fault	50.0%
Overproduction	Fault	100%
Pitch control deviation axis 3	Fault	100%
Safety chain	Fault	100%
Turbine OK	Turbine OK	99.45%
Weather downtime	Weather downtime	60.70%
Yaw runaway	Fault	87.50%

TABLE X
MODEL ANALYSIS ON TURBINE 17

Actual output	Anticipated output	Correctly identified cases
Centrifugal switch	Fault	100%
Gearbox oil over-temperature	Fault	0.00%
Maintenance downtime	Maintenance downtime	100%
Pitch overrun 0^0	Fault	100%
Power failure	Weather downtime	60.00%
Turbine OK	Turbine OK	99.27%
Weather downtime	Weather downtime	95.54%
Yaw runaway	Fault	100%

IV. CONCLUSION

In this paper, a methodology for predicting wind turbine states was presented. The proposed approach involved three key steps: turbine state abstraction, algorithm learning, and state prediction. In the first step, the initial wind turbine states were separated into classes using domain knowledge. To reduce the computational effort, data-mining algorithms were trained using a stratified data set. Turbine parameters such as the blade pitch angle, generator/gearbox speed, temperature hub, nacelle revolution, and current Phase C constitute the input to the prediction model. Among the selected data-mining algorithms, the random forest algorithm provided the most accurate results. Prediction models with up to 300-s horizons provided results with an accuracy in the range 78%–98%. The proposed model also identified various faults that occurred at wind turbines not included in the training data.

A month-long data set was available for this research. Future research will involve further analysis once additional data becomes available. New concepts will be researched to improve model robustness for identifying faults not reflected in the training data sets.

REFERENCES

[1] C. Hatch, "Improved wind turbine condition monitoring using acceleration enveloping," *Orbit*, pp. 58–61, 2004.

[2] A. Kusiak and W. Li, "The prediction and diagnosis of wind turbine faults," *Renew. Energy*, vol. 36, no. 1, pp. 16–23, 2011.

[3] C. Walford, Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs Sandia National Laboratories, Rep. SAND-2006-1100, Mar. 2006, 2006.

[4] R. Hyers, J. McGowan, K. Sullivan, J. Manwell, and B. Syrett, "Condition monitoring and prognosis of utility scale wind turbines," *Energy Mater.*, vol. 1, no. 3, pp. 187–203, 2006.

[5] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, and C. K. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 1, pp. 1–39, 2009.

[6] Y. Amirat, M. Benbouzid, B. Bensaker, and R. Wamkeue, "Condition monitoring and fault diagnosis in wind energy conversion systems: A review," in *Proc. 2007 IEEE Int. Electric Machines and Drives Conf.*, May 2007, vol. 2, pp. 1434–1439, 2007.

[7] P. Tavner, G. W. Bussel, and F. Spinato, "Machine and converter reliabilities in wind turbines," in *Proc. 3rd IET Int. Conf. Power Electronics Machines and Drives*, 2006, pp. 127–130.

[8] P. Caselitz and J. Giebhardt, "Rotor condition monitoring for improved operational safety of offshore wind energy converters," *J. Solar Energy Eng.*, vol. 127, no. 2, pp. 253–261, 2005.

[9] E. Becker and P. Posta, "Keeping the blades turning: Condition monitoring of wind turbine gears," *Refocus*, vol. 7, no. 2, pp. 26–32, 2006.

[10] "Managing the wind: Reducing kilowatt-hour costs with condition monitoring," *Refocus*, vol. 6, no. 3, pp. 48–51, 2005, Editorial.

[11] A. Kusiak and A. Verma, "A data-driven approach for monitoring blade pitch faults in wind turbines," *IEEE Trans. Sustain. Energy*, vol. 2, no. 1, pp. 87–96, Jan. 2011.

[12] P. Caselitz, J. Giebhardt, and M. Mevenkamp, "Application of condition monitoring systems in wind energy convertors," in *Proc. EWEC*, Dublin, 1997, pp. 1–4.

[13] A. Zaher and S. McArthur, "A multi-agent fault detection system for wind turbine defect recognition and diagnosis," in *Proc. IEEE Lausanne Powertech*, 2007, pp. 22–27.

[14] A. Kusiak, H. Y. Zheng, and Z. Song, "On-line monitoring of power curves," *Renew. Energy*, vol. 34, no. 6, pp. 1487–1493, 2009.

[15] A. Kusiak and A. Verma, "Prediction of status patterns of wind turbines: A data-mining approach," *Trans. ASME J. Solar Energy Eng.*, vol. 133, no. 1, pp. 1–10, 2011.

[16] A. Kusiak and Z. Zhang, "Adaptive control of a wind turbine with data mining and swarm intelligence," *IEEE Trans. Sustain. Energy*, vol. 2, no. 1, pp. 28–36, Jan. 2011.

[17] P. J. Tavner, G. J. W. Bussel, and F. Spinato, "Machine and converter reliabilities in wind turbines," in *Proc. Inst. Elect. Eng. 2nd Int. Conf. Power Electronics, Machine & Drives*, Dublin, Apr. 2006.

[18] P. J. Tavner and J. Xiang, "Wind turbine reliability, how does it compare with other embedded generation sources," in *Proc. Inst. Elect. Eng. RTDN Conf.*, London, U.K., Jan. 2005.

[19] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, 1997.

[20] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading: Addison-Wesley, 1989.

[21] A. Sbihi, "A best first search exact algorithm for the multiple-choice multidimensional knapsack Problem," *J. Combinatorial Optimization*, vol. 13, pp. 337–351, 2007.

[22] T. Kudo and Y. Matsumoto, "A boosting algorithm for classification of semi-structured text," in *Proc. EMNLP*, 2004, pp. 301–308.

[23] A. C. McCormick and A. K. Nandi, "Classification of rotating machine condition using artificial neural networks," *Proc. IMechE: Part C*, vol. 11, no. 6, pp. 439–450, 1997.

[24] V. Kecman, *Learning and Soft Computing-Support Vector Machines, Neural Networks, Fuzzy Logic Systems*. Cambridge, MA: MIT Press, 2001.

[25] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[26] G. V. Kass, "An exploratory technique for investigating large quantities of categorical data," *Appl. Stat.*, vol. 29, no. 2, pp. 119–127, 1980.

Andrew Kusiak (M'90) received the B.S. and M.S. degrees in engineering from the Warsaw University of Technology, Warsaw, Poland, in 1972 and 1974, respectively, and the Ph.D. degree in operations research from the Polish Academy of Sciences, Warsaw, in 1979.

He is currently a Professor in the Department of Mechanical and Industrial Engineering, University of Iowa, Iowa City. He speaks frequently at international meetings, conducts professional seminars, and does consultation for industrial corporations. He has served on the editorial boards of more than 40 journals. He is the author or coauthor of numerous books and technical papers in journals sponsored by professional societies, such as the Association for the Advancement of Artificial Intelligence, the American Society of Mechanical Engineers, etc. His current research interests include applications of computational intelligence in automation, wind and combustion energy, manufacturing, product development, and health-care.

Prof. Kusiak is the Institute of Industrial Engineers Fellow and the Editor-in-Chief of the *Journal of Intelligent Manufacturing*.

Anoop Verma (S'10) received the B.S. degree in manufacturing engineering from NIFFT, India, and the M.S. degree in mechanical engineering from the University of Cincinnati, Cincinnati, OH, in 2007 and 2009, respectively. At present he is working toward the Ph.D. degree in the Industrial Engineering Program, The University of Iowa, Iowa City.

He has researched areas such as machine loading, manufacturing scheduling, corporate memory, rapid prototyping, and information retrieval and published papers in various journals, including IEEE TRANSACTION ON SUSTAINABLE ENERGY, *Expert Systems with Applications*, *International Journal of Production Research*, *Engineering Applications of Artificial Intelligence*, *Journal of Intelligent Manufacturing*, etc. He reviewed papers for various journals, including the *International Journal of System Science*, *Information Science*, and *Journal of Intelligent Manufacturing*. His current doctoral research focuses on applications of data mining and computational intelligence in wind industry.