Using Domain Knowledge Features for Wind Turbine Diagnostics

R. Lily Hu[1], Kevin Leahy[2], Ioannis C. Konstantakopoulos[3], David M. Auslander[4], Costas J. Spanos[5], and Alice M. Agogino[6]
[1, 2, 4, 6]Department of Mechanical Engineering
University of California, Berkeley
[3, 5]Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
[2]Department of Civil and Environmental Engineering
University College Cork, Ireland
{[1]lhu, [2]leahyk, [3]ioanniskon, [4]leahyk, [5]spanos, [6]agogino}@berkeley.edu
[4]dma@me.berkeley.edu

***Abstract*—Maximising electricity production from wind requires improvement of wind turbine reliability. Component failures result in unscheduled or reactive maintenance on turbines which incurs significant downtime and, in turn, increases production cost, ultimately limiting the competitiveness of renewable energy. Thus, a critical task is the early detection of faults. To this end, we present a framework for fault detection using machine learning that uses Supervisory Control and Data Acquisition (SCADA) data from a large 3MW turbine, supplemented with features derived from this data that encapsulate expert knowledge about wind turbines. These new features are created using application domain knowledge that is general to large horizontal-axis wind turbines, including knowledge of the physical quantities measured by sensors, the approximate locations of the sensors, the time series behaviour of the system, and some statistics related to the interpretation of sensor measurements. We then use mRMR feature selection to select the most important of these features. The new feature set is used to train a support vector machine to detect faults. The classification performance using the new feature set is compared to performance using the original feature set. Use of the new feature set achieves an F1-score of 90%, an improvement of 27% compared to the original feature set.**

***Keywords*-Feature selection; domain knowledge; SCADA Data; Wind Turbine; Fault Detection; SVM; FDD; mRMR**

I. Introduction

Renewable energy generation is becoming increasingly important. For example, under the EU Renewable Energy Directive, Ireland needs to reach a binding target of sourcing 16% of its annual overall energy use from renewables by 2020. To reach this target, 40% of Ireland's electricity needs to come from renewable sources, for which the vast majority of this target will be met by wind energy [1].

However, wind turbines see highly irregular loads due to varied and turbulent wind conditions, and as a result, components can undergo high stress throughout their lifetime compared with other rotating machines [2]. Because of this, operations and maintenance account for up to 30% of the cost of generation of wind power [3].

The ability to remotely monitor component health is even more important in the wind industry than in other industries; wind turbines are often deployed to operate autonomously in remote sites so periodic visual inspections can be impractical. Unexpected failures on a wind turbine can be very expensive - corrective maintenance can take up a significant portion of a turbine's annual maintenance budget. Condition-based maintenance (CBM) is a strategy whereby the condition of equipment is actively monitored to detect impending or incipient faults, allowing an effective maintenance decision to be made as needed. This strategy can save up to 20-25% of maintenance costs vs. scheduled maintenance of wind turbines [4].

Machine learning approaches to CBM of wind turbines include principal component analysis and neural networks or other pattern recognition methods, usually applied to high resolution data from retrofitted vibration or oil particulate sensors. One such approach used a modified version of the classification method K nearest neighbors on 100Hz sampled data from a simulated benchmark model of a wind turbine [5].

Although Condition Monitoring Systems (CMSs) have been widely successful in other applications, CBM has not been taken up extensively by the wind industry, despite the supposed benefits [4]. A number of reasons exist for this [6], [7]. The capital cost of retrofitting sensors, as well as data collection and analysis can be quite high - upwards of €13,000 per turbine.

However, there already exist a number of sensors on modern turbines related to the Supervisory Control and Data Acquisition (SCADA) system. In recent years, there has been a concerted effort to apply condition monitoring (CM) techniques to wind turbines by analysing data collected by the SCADA system. SCADA data is typically recorded at 10-minute intervals to reduce transmitted data bandwidth and storage, and includes a plethora of measurements such as active and reactive power, generator current and voltages, wind speed, generator shaft speed, generator, gearbox and nacelle temperatures, and others [2]. By performing statistical analyses on various trends within this data, it is possible to detect when the turbine is entering a time of sub-optimal performance or if a fault is developing. This is all done without the added costs of retrofitting additional sensors to the turbine [6].

Previous work by the authors involved predicting wind turbine faults using classification methods applied to 10-minute SCADA data [8]. This paper aims to improve prediction scores by expanding the SCADA data features with features derived from wind turbine domain knowledge. These additional features are based on 1) an understanding of the physical values measured by the SCADA sensors, 2) the time series behaviour of the sensor measurements, and 3) statistical features. Feature selection methods are then applied to this expanded feature set to select the most important overall features and to validate whether the new features selected are more useful for prediction than the original ones. Next, support vector machines (SVMs) are trained using both the expanded feature set and the original SCADA data set to predict faults. The prediction results of the expanded feature set are then compared with the results of using only the SCADA data.

This paper is organised as follows: In section II, a description is given of the data. Section III describes the proposed approach. Section IV reports the experimental results. Section V summarises the conclusions.

Table I
10 MINUTE OPERATIONAL DATA

TimeStamp	**Wind Speed (avg.)** m/s	**Wind Speed (max.)** m/s	**Wind Speed (min.)** m/s	**Power (avg.)** kW	**Power (max.)** kW	**Power (min.)** kW	**Bearing Temp (avg.)** °C
09/06/2014 14:10:00	5.8	7.4	4.1	367	541	285	25
09/06/2014 14:20:00	5.7	7.1	4.1	378	490	246	25
09/06/2014 14:30:00	5.6	6.5	4.5	384	447	254	25
09/06/2014 14:40:00	5.8	7.5	3.9	426	530	318	25
09/06/2014 14:50:00	5.4	6.9	4.5	369	592	242	25

Table II
STATUS DATA

Timestamp	**Main Status**	**Sub Status**	**Status Text**
13/07/2014 13:06:23	0	0	Turbine in Operation
14/07/2014 18:12:02	62	3	Feeding Fault: Zero Crossing Several Inverters
14/07/2014 18:12:19	80	21	Excitation Error: Overvoltage DC-link
14/07/2014 18:22:07	0	1	Turbine Starting
14/07/2014 18:23:38	0	0	Turbine in Operation
16/07/2014 04:06:47	2	1	Lack of Wind: Wind Speed too Low

II. DATA

A. Data Source

Data is collected from a 3 MW direct-drive wind turbine near the coast in the South of Ireland for an 11-month period from May 2014 - April 2015. This wind turbine supplies power to a major biomedical devices manufacturing plant. The data is downloaded from the wind turbine's SCADA system and is comprised of three separate datasets: "operational" data, "status" data, and "warning" data.

The "operational" data has 54 features with a sampling resolution of every 10 minutes. These features include wind speed, ambient temperature, power characteristics such as real and reactive power, and temperatures of components in the wind turbine such as the generator bearing and rotor. For some of the quantities above, the features include the average, minimum and maximum over the 10 minute period. A sample of this data is shown in Table I.

The "status" and "warning" data is used to create labels and to process the "operational" data. The "status" and "warning" data are event logs, a sample of which is found in Table II. The "status" data records changes in the status of the wind turbine. The "warning" data is also called "information messages" and mostly corresponds to general information about the turbine.

B. Data Pre-Processing

The "operational" data is sorted by time stamp. Samples that are in-between the 10 minute sampling times are removed and samples with the same time stamp are averaged. The labels for the samples were determined based on the "status" data.

From the "warning" data, times corresponding to a single specific warning message (main warning code " *230 - Power Limitation (10h)*") are removed from the "operational" data set. This warning corresponds to slightly limited power output during nominal operation for one of a number of reasons, including turbine noise control during certain hours, an increase in internal temperatures on a hot day, or grid regulation. When this message is generated, there follows a 10-hour period where turbine power output may or may not be curtailed. Although considered a part of normal operation, for the purposes of this study it was decided to not include this data to give a clearer distinction for fault classification. It may be included in future work.

Class labels were created for the "operational" data using the "status" data. For this, a list of frequently occurring faults was made. For these faults, status messages with codes corresponding to the faults were selected. The start and end of these turbine states were used to match up the associated 10-minute operational data.

The wind turbine faults included in this data set are: *feeding faults*, *excitation errors*, *mains failure*, *aircooling malfunction* and *generator heating faults*. While the types of faults present in the data affect the features that are selected in the feature selection process, the methodology presented here can be applied to other faults. *Feeding faults* refer to faults in the power feeder cables of the turbine, *excitation errors* refer to problems with the generator excitation system, *mains failure* refers to problems with mains electricity supply *to* the turbine, *malfunction aircooling* refers to problems in the air circulation and internal temperature circulation in the turbine, and *generator heating faults* refer to the generator overheating. Thus, five types of faults are included in this data set. There are 213 occurrences of faults during the observation period, resulting in a total of 437 fault data samples.

The data is normalised and balanced class weights are used for training of machine learning models. There are a total of 54 original features.

III. Methodology

A. Overview

The "operational" data and corresponding features collected from the wind turbine are supplemented with additional features derived from the "operational" data based on domain knowledge. This process, shown in Figure 1, is novel in that it does not simply use the original sensor data treated as independent samples, but also includes time series characteristics and application domain knowledge; Furthermore, the original data is combined with this new knowledge in a simple form that is appropriate as input into numerous machine learning algorithms — a data matrix X of n rows of samples and m columns of features.

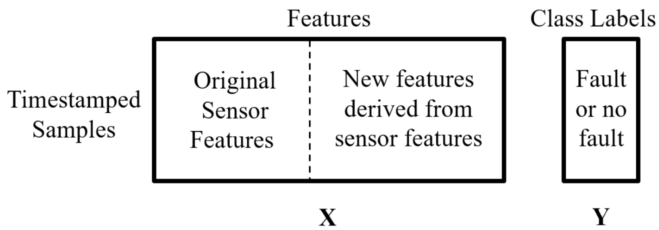

Figure 1. X and Y for machine learning models

B. Application Domain Knowledge Features

The original features in the "operational" data are supplemented with new features created from the original features using knowledge of wind turbines, namely, knowledge of the quantities that the original features correspond to; the location of where data for those features is collected; and an understanding of the operation of the wind turbine. For example, given knowledge that there are two nacelle ambient temperature features that correspond to the two ambient temperature sensors on the nacelle, it is expected that the two temperatures would have similar values. Thus, a new feature is the difference between the two nacelle temperatures, made possible because the original features have the same units. This and other instances of domain knowledge were acquired through interviewing wind turbine manufacturers and independent service provides, previous work experience in the wind industry, and basic science and engineering knowledge.

A summary of the new features derived from the original features using domain knowledge is shown in Table III, which represents 67 additional derived features.

Table III
Example Features from Knowledge of Wind Turbines

Average Of	Front and Rear Bearing Temperatures Rotor Temperatures Stator Temperatures ...
Difference Between	Max and Min of wind speed Max and Average of wind speed Min and Average of wind speed Front and Rear Bearing Temperature Nacelle Ambient Temperatures Generator Temperature and Nacelle Temperature ...
Ratio of	Average power to Available Power (from wind, technical reasons, force majeure reasons, force external reasons) ...

C. Time Series Features

The wind turbine is a physical system and it is known that the original features and derived features are physical quantities that form time series, hence timestamped "operational" data samples are not independent from one another. It is desirable to have the machine learning model work across time and to include correlation across sample columns. To that end, the "operational" data samples are converted to rolling time series representations of one hour using lagged features. The use of lagged features allows for the approximation of derivatives, an important aspect of physical systems. This is also similar to a finite impulse response filter (FIR), specifically a discrete time and digital FIR.

To represent the "operational" data as time series, lagged features — also called delays — for each original feature are created to include the data from time $t - 60min$ to t. For example, if the original feature "Nacelle Ambient Temp" is at time t and the sampling resolution is 10 minutes, one new feature is "Nacelle Ambient Temp at time $t-10min$", another new feature is "Nacelle Ambient Temp at time $t - 20min$", etc. all the way to "Nacelle Ambient Temp at time $t-60min$". This results in 324 time-lagged features.

D. Statistical Features

Statistical features are created from the original features. The first statistical feature is the 2-hr rolling average. For example, the rolling average of "Nacelle Ambient Temp" at sample time t is the average of "Nacelle Ambient Temp" between $t - 2hr$ to t. The rolling average is also a finite impulse response filter commonly called a boxcar filter. This has the benefit of suppressing high frequency noise. The second statistical feature is the 2-hr rolling standard deviation, together resulting in 108 statistical features.

E. Feature Selection

Given that the number of features are greatly increased, the mutual information based minimal-redundancy-maximal-relevance criterion (mRMR) feature selection method is used to find a subset of features useful for prediction of the faults. mRMR reduces redundancy in the features and selects those most relevant to prediction [9]. Both Mutual Information Difference (MID) and Quotient (MIQ) schemes are used.

F. Machine Learning

To see if the new derived features improve fault detection performance, support vector machines (SVMs) are trained using the new feature set and the original feature set for different numbers of features. The measured data is randomly separated into training and testing sets with 80% of the data in the training set and 20% of the data in the testing set. To train each SVM, a randomised grid search is performed

Table IV
TOP 10 FEATURES SELECTED BY MRMR UNDER THE MIQ AND MID SCHEME WITH FEATURE RANKINGS

Feature	MIQ	MID
Difference between P technical and P external	1	1
System 1 inverter 1 cabinet temp t-30min	2	NA
2hr mean of average blade angle A	3	NA
2hr stddev of spinner temp	4	NA
Difference between P technical and P majeure	4	NA
2hr mean of RTU ava Setpoint 1	6	2
2hr stddev of rear bearing temp	7	3
2hr mean of rotor temp 1	8	NA
Difference between avg Power and P from wind	9	NA
Average Nacel position including cable twisting t-60min	10	9
Difference between rotor temps	NA	4
Difference between P from wind and P technical	NA	5
Min windspeed t-60min	NA	6
Min windspeed t-20min	NA	7
Difference between nacelle ambient temps	NA	8
Difference between average and min rotation	NA	10

over hyperparameters using 10-fold cross validation to find the hyperparameters which yielded the highest F1 classification score. The hyperparameters that are searched over are C, γ, and the kernel. For each variation in the number of features, five SVMs are trained. From among the five SVMs, the SVM with the highest F1 score on the training set is used to predict on the testing set.

IV. RESULTS AND DISCUSSION

A. Selected Features

In the feature selection process, the new derived features are favoured more than the original features. The top ten features as selected by mRMR under the MIQ and MID schemes are shown in Table IV. All of the top ten features are new features. Furthermore, all the different types of new features — application domain knowledge features, time series features, and statistical features — are among the top ten features. There is some overlap between the features selected under the MIQ and MID schemes. Differences in rankings and selected features are expected because to select the features, the MIQ scheme uses a quotient while the MID scheme uses a difference.

Figure 2 plots the number of new derived features among the features selected by mRMR under the MIQ and MID schemes, along with their respective rankings. The dotted red line traces the expected curve if all the top features are new derived features. The actual curves follow this reference line closely. The deviations of the curves from the dotted reference line are the number of original features among the top selected features. This deviation is less than the number of original features selected if features were selected randomly from the new feature set.

Among the selected features, the derived features are often ranked higher than the features that they are derived from. The histogram in Figure 3 tallies the differences in rankings between the new derived features and their respective original features. To create Figure 3, the original features are binned together with their derived features. That is, features associated with the same original sensor reading are grouped together into a bin. The feature selection ranking of the first derived feature selected from that bin is compared to the feature selection ranking of the original feature by calculating the difference between the rankings. For example, if the standard deviation of "rear bearing temp" is the first derived feature selected from the "rear bearing temp" bin and has a ranking of 4, and the "rear bearing temp" feature has a ranking of 54, then the relative ranking is 50. Positive differences in ranking indicate that the derived feature is ranked higher than the original feature that it was derived from. The magnitude of the difference indicates how many rankings higher the derived feature is ranked compared to its original feature.

In Figure 3, all the ranking differences are positive, indicating that derived features are selected before the original features that they are derived from. Furthermore, the derived features are sometimes ranked substantially higher than the original features, with ranking differences in the hundreds.

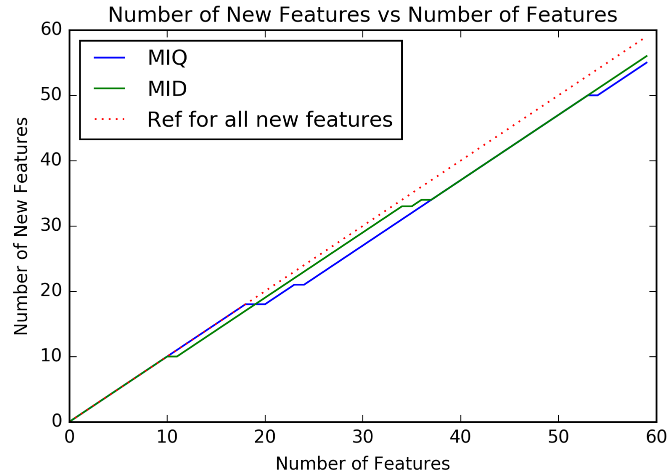

Figure 2. Number of new features among the top features with a dotted reference line indicating the expected line if all the features were new features

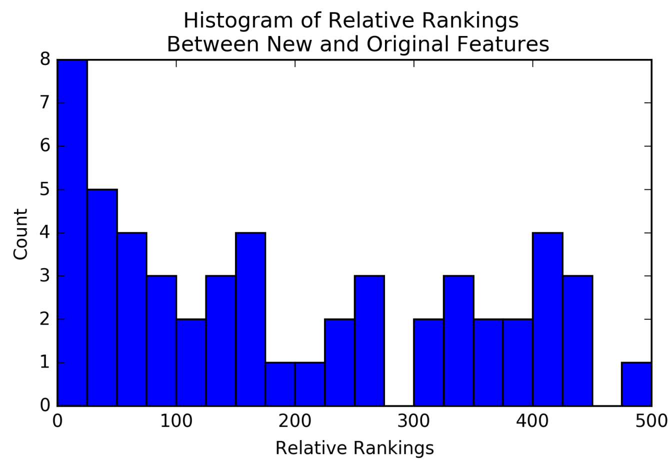

Figure 3. Histogram of relative ranking between original features and binned features

B. Model Prediction Results

Inclusion of the new derived features shows improvement in classification scores over the use of the original feature set for classification of faults and no-faults.

The classification results from training SVMs using different feature sets are shown in Figure 4. Figure 4 plots performance metrics on the y-axes for the 'fault' class of: F1 score on the first row of graphs, precision on the second row of graphs, and recall on the third row of graphs. The first column of graphs is the scores on the training set and the second column of graphs is the scores on the testing set. The x-axis is the number of features used in the SVM. Our ultimate goal is to save money and reduce computation time with fewer sensors and fewer features. From Figure 4, it is apparent that most of the benefit to prediction performance is already realized around 40 features, when the performance metrics plateaus, and thus the x-axis is not extended further.

The horizontal dotted red lines sketch out the classification scores for an SVM trained on the original features in the "operational" data set: an F1 score of 63%, precision of 77%, and recall of 53%. Values above the horizontal dotted red reference lines show classification performance better than performance achieved using the original features.

The vertical dotted red line traces the number of features in the original feature set; there are 54 features in the original feature set. Values to the left of the dotted red line indicate the use of fewer features than in the original feature set.

The solid blue line and dashed green lines plot the classification scores of SVMs trained using features from the new feature set as selected by mRMR under the MIQ scheme and the MID scheme respectively. Features selected using the MIQ scheme result in better classification scores than features selected using the MID scheme. This is observed in how the MIQ lines are higher than the MID lines.

Using new derived features selected by mRMR under the MIQ scheme, the same F1 score and precision as the original feature set is achieved using approximately two-thirds of the original number of features. Furthermore, improved recall is attained using as few as 20% of the original number of features.

When using a similar number of features as the original feature set, the F1 score improves by 19%, the precision by 7% and the recall by 28% on the test set. Even with 75% of the number of features, the F1 score is higher by 13%. The F1 score reaches 90% when 90 features from the new feature set are used, which is an improvement in F1 score of 27%. The selected features are likely to be dependent upon the wind turbine faults that are present in the dataset, and can be expected to vary with specific wind turbine faults.

These classification performance results demonstrate that the proposed procedure for incorporating expert domain knowledge can not only improve fault detection performance but can do so using fewer features. The use of fewer features enables savings from decreased data collection needs — such as from sensors, instrumentation, installation, networking, data quality, and data transfer needs — as well as avoiding problems that arise from machine learning on high dimensional data, such as large computation times and resource needs. The number of features used in an implementation of this fault detection system will thus be based on the desired trade-off between improved fault detection performance and data collection and computation costs.

V. Conclusion

The methodology proposed in this paper to incorporate expert knowledge by creating new features from existing sensor data enables higher classification scores and improved detection of faults while using fewer features — an improvement in F1 score of almost 20% while using a similar number of features, up to 27% with more features, and even an increase of 13% is possible while using only 75% as many features as in the original feature set. These improvements are seen using only basic, general knowledge of wind turbines that is not specific to the particular wind turbine installation.

The use of fewer features helps address problems that arise from a large number of dimensions, while saving data collection costs. Another benefit is that this approach allows for flexibility in the choice of machine learning algorithm because the data sets are in a format suitable for input into a wide variety of machine learning algorithms. Future work can explore the use of this methodology of incorporating domain knowledge using other machine learning algorithms.

Another avenue of investigation is to try other feature selection methods, particularly feature selection methods that take into account the correlation between features. Future research is also needed to explore how features selected by feature selection methods may be dependent on the faults present in the data. To investigate this behaviour, feature selection can be conducted on individual types of faults to investigate which features are good predictors for which faults. Additional research is also needed using more faults and different faults.

The proposed methodology can be applied to fault detection in other applications, such as other turbines, solar panels, and buildings. It would be interesting to compare and contrast performance and adapt this methodology across applications.

Acknowledgements

Kevin Leahy is supported by funding from MaREI, Marine and Renewable Energy Ireland and is a visiting doctoral student and researcher from the Department of Civil and Environmental Engineering from University College, Cork in Ireland.

Ioannis C. Konstantakopoulos is a Scholar of the Alexander S. Onassis Public Benefit Foundation.

This work is partially supported by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley

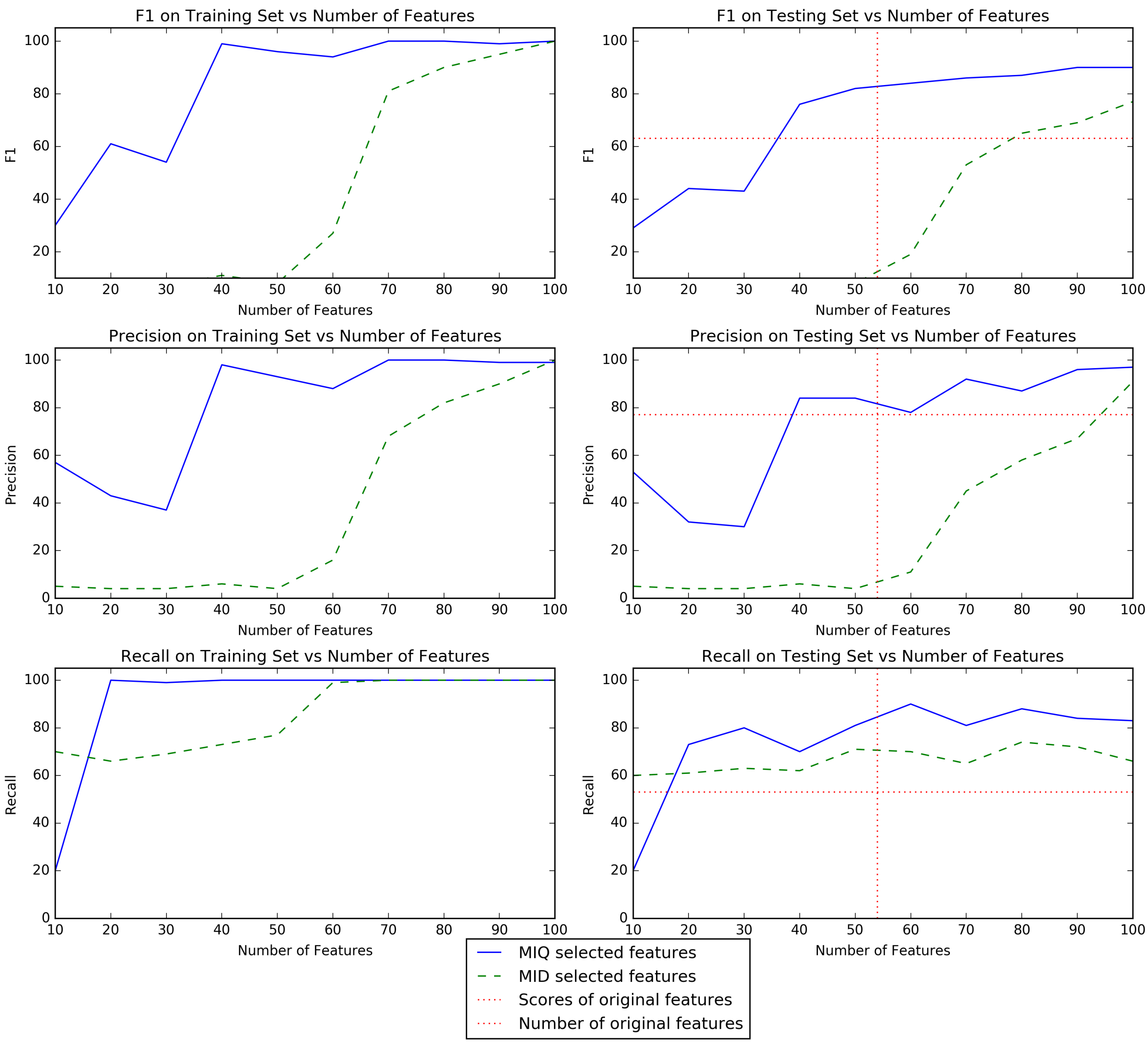

Figure 4. Scores on training and testing sets under the MIQ and MID scheme with reference scores and number of features of the original feature set

as a centre for intellectual excellence in research and education in Singapore.

Thanks to Andrew Lee for help with coding and access to computational resources. Thanks to Andrew Sabelhaus for help with accessing computational resources.

References

[1] S. Billington and P. Mohr, "The Value of Wind Energy to Ireland," 2014.

[2] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, sep 2009.

[3] European Wind Energy Association (EWEA), "The Economics of Wind Energy," Tech. Rep., 2009.

[4] J. L. Godwin and P. Matthews, "Classification and Detection of Wind Turbine Pitch Faults Through SCADA Data Analysis," *International Journal of Prognostics and Health Management*, vol. 4, p. 11, 2013.

[5] H. Toubakh, M. Sayed-Mouchaweh, and E. Duviella, "Advanced pattern recognition approach for fault diagnosis of wind turbines," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 2. IEEE, 2013, pp. 368–373.

[6] W. Yang, P. J. Tavner, C. J. Crabtree, Y. Feng, and Y. Qiu, "Wind turbine condition monitoring: Technical and commercial challenges," *Wind Energy*, vol. 17, no. 5, pp. 673–693, 2014.

[7] W. Yang, R. Court, and J. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis," *Renewable Energy*, vol. 53, pp. 365–376, may 2013.

[8] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, and A. M. Agogino, "Diagnosing Wind Turbine Faults Using Machine Learning Techniques Applied to Operational Data," *IEEE International Conference on Prognostics and Health Management*, 2016.

[9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.