A PCA and Two-Stage Bayesian Sensor Fusion Approach for Diagnosing Electrical and Mechanical Faults in Induction Motors

Anna Stief, James R. Ottewill, *Member, IEEE*, Jerzy Baranowski, *Member, IEEE*, and Michal Orkisz, *Senior Member, IEEE*

***Abstract*—Induction motors are widely used in industrial plants for critical operations. Stator faults, bearing faults, or rotor faults can lead to unplanned downtime with associated cost and safety implications. Different sensors may be used to monitor the health state of induction motors with each sensor typically being better suited for diagnosing different faults. Condition monitoring approaches that fuse data from multiple sensors have the potential to diagnose a greater number of faults. In this paper, a sensor fusion approach based on the combination of a two-stage Bayesian method and principal component analysis (PCA) is proposed for diagnosing both electrical and mechanical faults in induction motors. Acoustic, electric, and vibration signals are gathered from motors operating under different loading conditions and health states. The inclusion of the PCA step ensures robustness to varying loading conditions. The obtained results highlight that the proposed method performs better than the equivalent single-stage or feature-based Bayesian methods.**

***Index Terms*—Bayes methods, condition monitoring, fault detection, induction motors, principal component analysis (PCA), sensor fusion.**

I. Introduction

INDUCTION motors are widely used in industrial plants for critical operations, where a failure could result in a partial or complete shutdown of the production process. Unplanned maintenance, downtime, or replacements can result in high costs and, furthermore, critical failures can have serious safety implications. Induction motor faults may be categorized as electrical related, mechanical related, or environmental related [1]. The range of possible faults is numerous, with stator, bearing, and rotor faults being the most prevalent [2]–[4]. These faults will impact the mechanical, magnetic, and electrical characteristics of the induction motor in different ways. As a result, the optimal sensor type for diagnosing one type of fault mode may not be the same as the optimal sensor to diagnose another fault mode. It has previously been shown that specific induction motor faults can be diagnosed using different sensors [5]–[8]. Vibration, acoustic, and electric signals are among the most commonly used sensor types for rotor and stator faults detection, however some sensors are more suitable for detecting specific faults than others [7], [8]. Nandi [5] observed that acoustic and vibration signals are the most sensitive for bearing fault detection, whereas electric signals are more sensitive to broken rotor bar faults. It has recently been shown that acoustic signals are suitable for bearing, stator, and rotor fault diagnostics of single-phase and three-phase induction motors [9], [10]. Additionally, sensors that are responsive to a specific fault can also provide information about other faults [6]. Hence, a condition-monitoring system that fuses information obtained from multiple sensor types can ensure that a comprehensive range of fault modes may potentially be detected quickly and accurately.

Various condition-monitoring methods that aim to increase the accuracy and robustness of fault detection via sensor fusion have been reported. In [11], neural networks were used to fuse vibration and current signals in order to diagnose mechanical and electrical faults. It was shown that these signal types are complementary to one another and that their fusion using the Dempster–Shafer theory at the decision level increases the accuracy of the classification. A *K*-nearest neighbor classifier was applied in [12] using an accelerometer and load signals in order to diagnose bearing faults, showing that, whereas load signals are more useful in distinguishing healthy bearings from faulty ones and accelerometer signals are better at detecting the location of the fault, the best performance was achieved when the two signals were fused together. In [13], vibration and acoustic signals were fused using the Dempster–Shafer theory at the decision level to diagnose faults in planetary gearboxes, with the fusion resulting in more precise diagnostics along with reduced false and missed alarm rates. In [14], vibration, acoustic, and oil debris signals were fused at the feature level to diagnose faults in gears with principal component analysis (PCA) and independent component analysis. In each aforementioned case, the sensor fusion proved to increase the accuracy, robustness,

Manuscript received June 6, 2018; revised September 24, 2018, October 22, 2018, and November 27, 2018; accepted December 6, 2018. Date of publication January 14, 2019; date of current version July 31, 2019. This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement 675215. *(Corresponding author: Anna Stief.)*

A. Stief, J. R. Ottewill, and M. Orkisz are with the ABB Corporate Research Center, 31-038 Krakow, Poland (e-mail: anna.stief@pl.abb.com; james.ottewill@pl.abb.com; michal.orkisz@pl.abb.com).

J. Baranowski is with the Department of Automatic Control and Robotics, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, 31-038 Krakow, Poland (e-mail: jb@agh.edu.pl).

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/TIE.2019.2891453

and missed or false alarm rate of the system. Sensor fusion can be implemented at the data level, the feature level, and at the decision level. The decision on the abstraction level depends on the information carried by the different signals. If the signal types are significantly different and carry complementary information, it is advised to use decision-level fusion [11], [15].

A typical challenge encountered when creating decision-level fusion algorithms is that there are often a large number of features relative to the number of observations. These features can be highly correlated, which ultimately can bias the results of the fault detection algorithm. A common method to reduce the correlation and the dimensionality of the features is PCA [16], [17]. For example, in [18], the dimensionality of features extracted from vibration and current signals was reduced by PCA before applying genetic algorithms and an artificial neural network for classifying faults in an induction motor. It was found that the performance of the fault classifier was improved by adding PCA as a feature preprocessing step. In [19], several feature reduction and transformation methods including neighborhood component analysis, linear discriminant analysis (LDA), locally linear coordination, and PCA were compared with maximally collapsing metric learning for multiple bearing fault diagnosis in induction motors with particular focus given to the dimensionality reduction aspect. Feature reduction is also found in multistage frameworks for the induction motor diagnosis, for example, a recent work [20] applied PCA, LDA, a genetic algorithm, and the Fisher score in a hybrid strategy to obtain a reduced and optimized feature set from vibration signals.

Another regularly observed fault detection problem is the varying operating conditions of the machines, which can originate from a change in the load or environmental conditions. In [21], it was concluded that the prediction performance of a support vector machine (SVM) based fault detection algorithm for mechanical and electrical fault detections in induction motors is load dependent. Different severities of stator faults were monitored in induction motors under changing load torque and supply voltage unbalances in [22], finding that the performance of a multiagent system and neural estimator depends on the severity of the fault. Diagnostics and prognostics methods of rotating machinery were reviewed in [23], highlighting the operating condition dependence of algorithms as an existing but an understudied area.

Bayesian inference has been described as a suitable method for fault detection and fault classification in condition-monitoring systems [23], [24]. Recently, Jaramillo *et al.* [25] proposed a two-stage Bayesian inference approach to monitor the condition of a system composed of several subsystems. The first stage of the sensor fusion takes place at the subsystem level, whereas the second stage fuses the result of the first stage at the decision level in order to determine the health state of the whole system. The method was efficient in diagnosing faults in complex systems composed of interacting components.

Existing two-stage Bayesian sensor fusion frameworks described in the literature [25], [26] typically set alarm thresholds according to the probability distributions of features and control limits. Properly tuning alarm thresholds can be challenging, particularly when there are a large number of features in the data set, or when the thresholds themselves might optimally be described as a function of other parameters (e.g., operating conditions).

This paper is an extension of the previous work in which a two-stage Bayesian sensor fusion method was applied to the diagnosis of mechanical faults in induction motors [26]. It was shown that, by fusing independent diagnoses of different sensor types at the decision level, the false and missed alarm rates of a fault classification algorithm could be significantly reduced. In [26], simple linear models of expected feature values relative to load values were applied to account for the load dependence of features. Such an approach limits the generality of the solution as the loading of the system is also required as an input to the algorithm during training and testing. It was also observed that the features used for training the Naïve Bayes classifier were highly correlated. As previously noted, such correlations between features can potentially bias the fault detection algorithm toward certain diagnoses.

In this paper, a two-stage (local and global) Bayesian method combined with PCA is proposed as a method for diagnosing not only mechanical but also electrical faults in induction motors operating under varying load and environmental conditions. Stator, rotor, and bearing faults are all considered. Features are extracted from acoustic, electric, and vibration signals recorded from an experimental system. PCA is used to remove the correlations that are present in the extracted features and reduce the influence of load conditions. At the local Bayesian stage, principal components of the features are fused with a Gaussian Naïve Bayes (GNB) classifier. At the global Bayesian stage, the results of the local stages are fused in order to create a final diagnosis. The generality of the algorithm is investigated by omitting data recorded at selected operating and environmental conditions from the training set and subsequently testing the trained model using the omitted data.

The novelties of this paper are as follows.

1) A two-stage Bayesian sensor fusion approach is extended by integrating PCA and GNB classifiers into the framework.
2) It is known that many fault indicators are dependent on loading conditions. By incorporating a multivariate statistical approach into the analysis, the correlations between operating conditions and feature level are accounted for. It is shown that the resulting method is able to accurately diagnose faults even for loading conditions not present in the training set.
3) In this paper, additional data addressing stator faults with varying severity are included into the analysis. This data is used to illustrate how, by fusing the different signals, it is possible to achieve a holistic monitoring solution that both provide greater coverage and greater monitoring accuracy compared to considering each sensor independently.
4) Through the addition of PCA and the GNB classifier, the approach introduced in this paper does not require monitoring thresholds to be defined, as the posterior fault class probabilities are directly calculated.

This paper is organized as follows. In Section II, the methods are introduced. In Section III, the experimental data are described, which were used for the validation of the methods. Section IV describes the implementation of the methods using the experimental data. The results of the proposed fault diagnosis method are presented in Section V with a discussion in Section VI. Finally, in Section VII, conclusions are drawn, pointing out the advantages, limitations of the method, and possible future work.

II. Methods

A. Principal Component Analysis

PCA is a well-established method for feature extraction, dimensionality reduction, data compression, and data visualization [27]. It is a common problem in data analysis that the features or attributes of the observation data are highly correlated. PCA transforms the correlated features to a linear space where the transformed features are uncorrelated and are ordered in a way that the first features retain most of the variation in the data. Singular value decomposition or eigenvalue decomposition (EIG) are popular algorithms for performing PCA. Here, SVD is considered, as it is numerically more robust when matrices are either singular or numerically very close to singular. Furthermore, SVD directly provides the required scores and loadings. If $\mathbf{X}$ is an $n \times m$ matrix with rank r, with n observations and m features, SVD is defined as follows:

$$\mathbf{X} = \mathbf{ULA}^{\mathrm{T}} \tag{1}$$

where $\mathbf{U}$ is an $n \times r$ orthonormal matrix, $\mathbf{L}$ is an $r \times r$ diagonal matrix, and $\mathbf{A}$ is an $m \times r$ orthonormal matrix. $\mathbf{UL}$ is an $n \times r$ matrix, containing the transformed uncorrelated features in the principal component space, usually referenced as scores. $\mathbf{A}$ contains the principal components, sometimes called loadings. For further information on PCA and SVD, readers are guided to [27].

B. GNB Classifiers

A GNB classifier is a probabilistic classifier, which assumes conditional independence between data that are distributed according to a Gaussian distribution. The classifier uses the Bayes theorem to calculate the posterior probabilities that an observation $\mathbf{x}_t = \{x_1, x_2, \ldots, x_m\}$ belongs to class c_i out of classes $\mathbf{C} = \{c_1, c_2, \ldots, c_p\}$ in the following way:

$$P(c_i|\mathbf{x}_t) = \frac{P(c_i) \cdot \prod_{j=1}^{m} P(x_j|c_i)}{\sum_{k=1}^{n} P(c_k) \cdot \prod_{j=1}^{m} P(x_j|c_k)} \tag{2}$$

where $P(c_i)$ is the prior probability of an observation belonging to class c_i. The classifier learns the $P(x_j|c_i)$ conditional probabilities that a given feature value x_j belongs to class c_i from a training dataset. By assuming a Gaussian distribution of the features, the conditional probabilities may be obtained using the values of mean and standard deviation of the labeled training data for each class as follows:

$$P(x_j|c_i(\mu_{i,j}, \sigma_{i,j})) = \frac{1}{\sigma_{i,j}\sqrt{2\pi}} \cdot e^{\frac{-(x_j - \mu_{i,j})^2}{2\sigma_{i,j}^2}}. \tag{3}$$

Once the posterior probabilities are calculated for all of the classes, the observation $\mathbf{x}_t$ will be classified into the class that has the highest posterior probability. Equation (2) can be simplified by omitting the normalization factor in the denominator, as only the index of the maximum *a posteriori* (MAP) class is important for the classification

$$P(c_i|\mathbf{x}_t) \propto P(c_i) \cdot \prod_{j=1}^{m} P(x_j|c_i) \tag{4}$$

$$c_{\text{predicted}} = \arg\max \left\{ P(\overline{C}|\overline{\mathbf{x}_t}) \right\}. \tag{5}$$

For further reference regarding GNB classifiers, readers are guided to, for example, [28]–[30].

C. PCA and Two-Stage Bayesian Sensor Fusion

The proposed two-stage Bayesian sensor fusion method combined with PCA is an extension of a previous work [26]. In this paper, the algorithm is updated to include a preprocessing PCA step. PCA was selected as it is able to mitigate feature correlation that can bias the likelihood calculations. It is a linear method that yields a reduced and uncorrelated feature set. Instead of the original features, uncorrelated principal components are fused using a GNB classifier. The number of principal components considered for each signal type is calculated using the validation set in a way that the performance of the algorithm is maximized while the false and missed alarm rates are reduced, using the detection accuracy as an optimization parameter. The method retains the structure of the global fusion stage on the decision level, as described in [26]. The advantage of applying the GNB classifier at the local stage is that there is no need to determine alarm thresholds and confidence intervals, as the GNB classifier calculates the fault class probabilities directly.

D. Description of the Local Stage

The proposed algorithm is suited for condition-monitoring problems where N different sensors provide measurement data for the determination of the health state of the system. For training, the algorithm requires data that has been labeled with M fault conditions. If there is a test set available, the data has to be split into two separate datasets for training: the training set and the validation set. The training set will be used for the training of the GNB classifiers at the local stage, whereas the validation set will produce the confusion matrices for the different sensor types at the global fusion stage.

Once the data are cleaned and selected features are extracted, the features are split by sensor type. At this stage, the training set takes the form of an $n \times m$ matrix, where n is the number of observations and m is the number of features. The $\mu_{Ai,Sj}$ means and $\sigma_{Ai,Sj}$ standard deviations are calculated for each A_i feature and S_j sensor type. A normalization step transforms the features such that the means are 0 and the standard deviations are 1. PCA calculates the SC_{Sj} scores and LO_{Sj} loadings for each sensor type. The scores, which might also be considered as the new "features," are uncorrelated. The LO_{Sj} loadings are calculated using the whole training set containing both healthy and faulty data. To calculate the conditional probabilities of

the GNB according to (3), the $\mu_{Ai,Sj,Ck}$ means and $\sigma_{Ai,Sj,Ck}$ standard deviations of the principal components are calculated for each C_k fault type in the labeled data.

Next, the validation set is used in both to find the optimal number of principal components and to calculate the confusion matrices using $\mu_{Ai,Sj}$, $\sigma_{Ai,Sj}$, $\mu_{Ai,Sj,Ck}$, $\sigma_{Ai,Sj,Ck}$, and LO_{Sj} from the training data. The features in the validation set are normalized using $\mu_{Ai,Sj}$ and $\sigma_{Ai,Sj}$. The normalized features are transformed to the principal components space using the LO_{Sj} loadings. To find the number of principal components for each S_j sensor type, an iterative step is considered as follows.

1) The first *i* principal components are used as features, calculating the posterior probabilities and class predictions for each observation in the validation set using (3)–(5).
2) Count the correct predictions and save it for *i*.

Once the iteration has finished, the value of *i* resulting in the highest number of correct predictions is chosen for the number of principal components used to calculate the predictions for each observation in the validation set.

E. Description of the Global Stage

The prediction counts for each fault type are organized in an $M \times M$ global confusion matrix $\mathbf{G}_{Si}$ for each sensor type S_i where the rows represent the actual condition, the columns represent the diagnosed condition, and the prediction counts by rows are divided by the total number of actual conditions for the fault type. The matrix elements can be interpreted as $P(F_i|F_j)$ conditional probabilities; given that the algorithm predicted F_j, what is the probability that the actual fault condition is F_i? The $P(F_i|F_i)$ probabilities, located along the diagonal of the confusion matrix for each sensor type, represent the probability that the sensor diagnosed the corresponding fault correctly

$$\mathbf{G}_{S_i} = \begin{bmatrix} P(F_1|F_1) & \dots & P(F_1|F_M) \\ \dots & P(F_i|F_i) & \dots \\ P(F_M|F_1) & \dots & P(F_M|F_M) \end{bmatrix}. \quad (6)$$

The test set is separate from the training set and is divided by sensor type into *N* sets, with observations in rows and features in columns. The test set is normalized and the GNB classifier is calculated with the optimized number of principal components.

The fault class predictions of the GNB classifier for an observation are fused by (7) and (8) using the appropriate columns from the global confusion matrices for each sensor type. $P(c_i)$ represents *a priori* knowledge; if no prior distribution is available, a uniform distribution is supposed. If the fault class predicted by S_1 is F_i and fault class predicted by S_M is F_j, then columns have to be selected in the following way from the corresponding confusion matrices:

$$\mathbf{G}_{S_1,F_i} = \begin{bmatrix} P(F_1|F_i) \\ \dots \\ P(F_M|F_i) \end{bmatrix}, \dots, \mathbf{G}_{S_M,F_j} = \begin{bmatrix} P(F_1|F_j) \\ \dots \\ P(F_M|F_j) \end{bmatrix} \quad (7)$$

$$c_{\text{predicted}} = \arg\max \left\{ P(c_i) \cdot \prod_{i=1,j=1}^{M,N} \mathbf{G}_{S_i,F_j} \right\}. \quad (8)$$

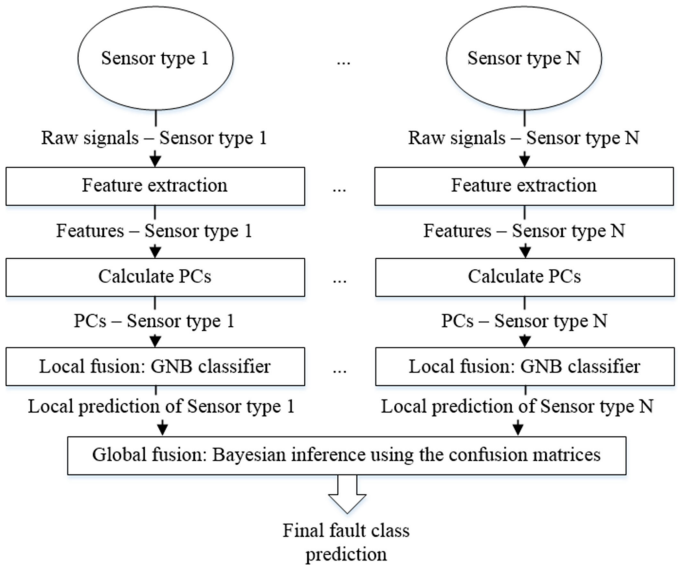

Fig. 1. Structure of the PCA and two-stage Bayesian algorithm.

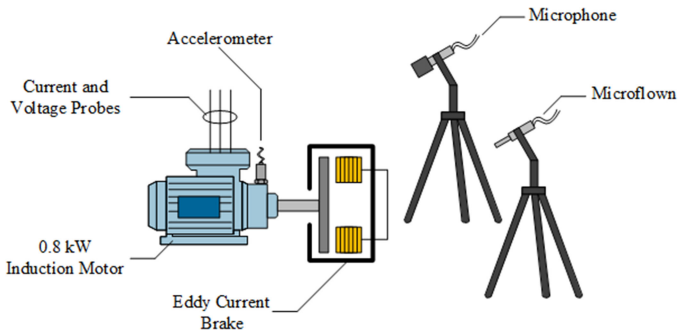

Fig. 2. Schematic of the experimental system.

For each fault class, the output of the global fusion step is *a posterior* probability giving the likelihood of that fault class being present in the system. For the purposes of evaluating the performance of the algorithm, we consider the final prediction as being the fault class that has the highest posterior probability after the global fusion step (8).

F. Testing an Observation

The overall flow diagram of the proposed two-stage Bayesian sensor fusion method for testing an observation is shown in Fig. 1. At the local stage, each type of sensor is handled separately. For a given sensor type, features are fused in order to obtain a prediction of the most likely health state of the system, given the data recorded by that sensor type. At the global stage, the predictions of the most likely health state for each sensor type are fused using the global confusion matrix to create the global diagnosis result.

III. Experimental Data

The measurement set up for the experiment is shown in Fig. 2. Experimental data were collected from three identical induction motors, differing only in terms of health state: one motor was healthy, one had two broken rotor bars, and one had an outer

raceway fault in a bearing. It was also possible to seed stator faults into the nominally healthy motor, as described in [31]. The test motors were 0.8 kW, four-pole SZJKe 14a induction motors manufactured by TAMEL with a nominal rotor speed of 1400 r/min. The nominal values of voltage, current, rated torque, and power factor for these motors were 380 V, 2.2 A, 5.45 N·m, and 0.74, respectively. The motor had a Y winding configuration with 4 coils per phase, 22 rotor bars, and 24 stator slots. The rotor inertia was 0.0025 kg·m^2 and the motor bearings were SKF type 6304 ZZ CXSQ. An eddy current brake was used to load the motor. The measurements were conducted at steady-state operation under different loading conditions. For each fault case between three and five loading conditions were tested, resulting in stator currents of 68%, 81%, 90%, 100%, and 113% of nominal values. Measurements were recorded both with and without background noise generated by a separate shaker. Datasets for eight different health conditions were recorded, denoted as F_0–F_7, as follows:

1) F_0—Healthy motor;
2) F_1—Stator fault: Phase one bypassed in the first phase;
3) F_2—Stator fault: Phase one bypassed in half of the first phase;
4) F_3—Stator fault: Phase–phase short circuit;
5) F_4—Stator fault: Phase–phase short circuit with an offset point;
6) F_5—Stator fault: Break of half of the phase one;
7) F_6—Rotor fault: Two broken rotor bars;
8) F_7—Bearing fault: Outer raceway defect.

The tested motor was rewound in such a way that instead of coils for a given phase being directly connected to one another, the individual coils were connected to a switchboard allowing the winding configuration to be quickly changed. Furthermore, in six coils, special taps were created in order to allow different short circuits to be seeded. Such a configuration allows various stator faults to be seeded, as was investigated in [29] for the same SZJKe 14a induction motor. For F_1 and F_2, the first phase was bypassed by a 15 Ω resistance causing a short circuit in the first phase winding. For F_3 and F_4, a short circuit of two stator phases in the taps connected in the middle of the first coils was seeded by adding a 115 Ω resistance. In the case of F_5, part of the coil was not connected causing asymmetry in the winding, so that the current did not flow through a part of the winding. The two broken rotor bars (F_6) were located next to one another. The bearing fault (F_7) was caused by an incision through the outer ring of the bearing.

Acoustic, electric, and vibration signals were collected using five different sensor types. Three G.R.A.S. 46AE microphones were used to measure the sound pressure levels. A Model USP regular three-dimensional Sound Intensity Microflown probe was also used to collect acoustic signals from the motors. The probe provided four measurement signals, three particle velocity signals in three orthogonal directions and a sound pressure signal. The vibration signals were measured by a three-axis PCB ICP accelerometer Model No. 356B18 and a one-axis PCB ICP accelerometer Model No. 353B32, providing four signals in total in unit g. The three phase voltages were measured by LV 25-P voltage transducers providing signals directly for analysis

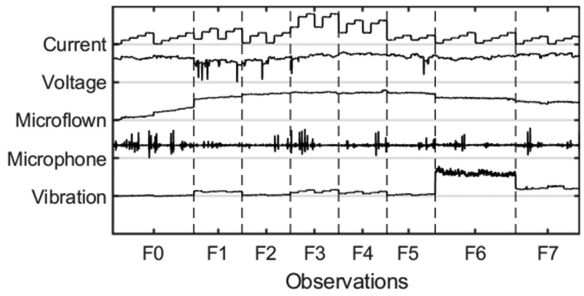

Fig. 3. Relative rms values of 5 different signal types extracted from 0.5-s measurement windows, for all observations through the 58 datasets.

of voltage characteristics. The motor currents were measured by LTS-6NP and LEM HY 5-P current transducers. The following signals were collected using a 16-channel LMS Scada Mobile System: 4 microflown signals, 3 microphone signals, 2 current signals, 4 vibration signals, and 3 voltage signals. Data were collected with a 51.2-kHz sampling rate to capture all frequencies of interest with 30 s of data being recorded for each configuration to capture a sufficiently long steady-state periods for analysis. 58 datasets were obtained: one for each tested loading condition, both with and without additional background noise. The same background noise was applied over the tests. The microflown axis X probe has measured an average 47.26 m/s particle velocity with no noise, whereas it has measured an average 88.69 m/s particle velocity with noise for the healthy motor under nominal load.

IV. Implementation of the Method

The 58 datasets were split into 0.5-s observations resulting in 60 observations for 1 dataset and 3480 observations in total. For each signal, and for each 0.5-s observation, the following time-domain features were extracted: root mean square (rms), skewness, kurtosis, maximum peak, peak-to-peak, and crest factor.

Features were also extracted from both the amplitude spectrum and the envelope spectrum of the signal: the frequency center, spectrum area, the amplitude of the components at the first two harmonics of the supply frequency (50, 100), the first three harmonics of the rotation speed (1×, 2×, 3×), the amplitude ratios (2×/1×, 3×/1×), and the amplitude at the sidebands of the supply frequency (50 Hz ± 2 × slip, 50 Hz ± rotation speed). The 0.5-s window length provided a 2-Hz spectral resolution. While no windowing functions were applied in the calculation of the spectra, edge effects were found to be minimal. In total, 30 features were extracted for the 16 signals, resulting in 480 features in total. These time- and frequency-domain features are standard metrics, commonly used for the condition monitoring of induction motors [11], [21], [32]. It should be noted that for all signal types, all of the above-mentioned feature types were extracted. No additional feature selection approaches were applied.

Fig. 3 shows the relative rms values of five different signal types extracted from 0.5-s measurement windows, for all observations through the 58 datasets. It may be observed that the

TABLE I
TEST CASE A: RANDOM SPLIT

		Diagnosed condition							
		F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7
Actual condition	F_0	0.94	0.00	0.06	0.00	0.00	0.00	0.00	0.00
	F_1	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	F_2	0.02	0.00	0.98	0.00	0.00	0.00	0.00	0.00
	F_3	0.00	0.00	0.00	0.98	0.02	0.00	0.00	0.00
	F_4	0.00	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	F_5	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	F_6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	F_7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

TABLE II
TEST CASE B: LOWEST LOAD AND NO NOISE

		Diagnosed condition							
		F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7
Actual condition	F_0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	F_1	0.02	0.97	0.00	0.00	0.00	0.01	0.00	0.00
	F_2	0.22	0.07	0.57	0.00	0.00	0.14	0.00	0.00
	F_3	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	F_4	0.00	0.00	0.00	0.00	0.80	0.20	0.00	0.00
	F_5	0.02	0.07	0.03	0.00	0.13	0.75	0.00	0.00
	F_6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	F_7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

sensors reacted to the fault modes and loading conditions in different ways. For example, the rms current is increased for stator fault modes F_3 and F_4, whereas the rms vibration did not significantly react. Conversely, in the case of the rotor fault F_6, the vibration signal exhibited increased rms values, whereas the rms current did not show significant increases. This further illustrates that different faults are more easily diagnosed by different sensors. The 480 features of the 3480 observations were grouped by signal types into five groups, namely vibration features, current features, microflown features, microphone features, and voltage features. The data were then split into a training set, a validation set, and a test set, in the same way for the five signal types. The division is described in Section V. The training sets were used to train the local stage, the validation sets were used to calculate the global confusion matrices for the global fusion stage, and finally, the test sets were used to test the performance of the algorithm. All analyses were conducted in MATLAB.

V. RESULTS

In order to illustrate the performance of the described algorithm with respect to different loading and environmental noise conditions, the experimental data were divided into different training, validation, and test sets. In Test Case A, a random split was applied. In Test Cases B and C, eight entire datasets (one from each fault case) were included in the test set with no datasets from experiments conducted at this loading condition being considered in the training or validation sets. In Test Case B, the lowest load datasets with no background noise are the test set. In Test Case D, the highest load datasets with background noise are the test set. The aim of testing different divisions for testing, validation, and training is to observe the performance of the algorithm under different operating conditions, particularly under loading conditions that were not considered during model training.

A. Test Case A: Random Split

Test Case A was used to evaluate the overall performance of the algorithm. The total 3480 observations were randomly split into training set, validation set, and test set with a respective ratio of 60-20-20%. The random split was applied 100 times and the averaged results are shown in Table I. The columns represent the conditions diagnosed by the algorithm, whereas the rows represent the actual fault conditions of the motors. The healthy motor was correctly diagnosed in 94% of the cases with a 6% false alarm rate in case of F_2 stator fault. Missed alarms are present for F_2, however it is only 2%. F_2 is the least severe fault among the seven seeded faults, which explains this behavior. The successful detection rate is above 98% for all fault cases, with 100% success rate for F_1, F_5, F_6, and F_7. Among the stator faults, the following scenario can be observed: F_3 and F_4 are sometimes misdiagnosed as each other, as they are the variations of the same fault: F_3 is the phase–phase short-circuit, whereas F_4 is the phase–phase short circuit with an offset point. To give an overall measure of the test accuracy, the F_1 score is calculated to be 99.32%.

B. Test Case B: Lowest Load and No Noise

In *Test Case* B, the test set was formed of data taken from the lowest loading conditions, with no datasets from experiments conducted at this loading condition being considered in the training or validation sets. The aim was to test the performance of the algorithm under load conditions that are lower than those contained within the training and validation sets. The results are shown in Table II. The accuracy of the algorithm was 100% when diagnosing the healthy condition (F_0); there were no false alarms. When diagnosing broken rotor bars and bearing faults (F_6 and F_7), the algorithm performed with 100% accuracy. However, the performance for the stator faults needs further analysis: while faults F_1 and F_3 are diagnosed with the success rates of 97% and 100%, faults F_2, F_4, and F_5 were identified less reliably. The algorithm was able to diagnose the F_2 stator fault in only 57% of the cases. In 43% of the cases, the algorithm misdiagnosed F_2, either as healthy or as the other similar stator faults F_1 and F_5. This was because F_2, as the least severe fault, was the most difficult to diagnose. The algorithm was also unable to distinguish between fault modes F_4 and F_5, in 20% and 13% of the cases, respectively. F_5 was also mistakenly diagnosed as other stator faults phase one bypassed in 10% of the cases. This result indicates that in the case of loading conditions lower than those seen in the training datasets, the algorithm can accurately determine the type of fault, however it is unable to accurately ascertain the severity of the fault.

C. Test Case C: Highest Load With Noise

Test Case C used datasets recorded for the highest loading conditions with background noise as the test set, with no data from this loading condition being considered in the training.

TABLE III
TEST CASE C: HIGHEST LOAD WITH NOISE

		Diagnosed condition							
		F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7
Actual condition	F_0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	F_1	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	F_2	0.02	0.00	0.98	0.00	0.00	0.00	0.00	0.00
	F_3	0.00	0.00	0.00	0.92	0.08	0.00	0.00	0.00
	F_4	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	F_5	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	F_6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	F_7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

TABLE IV
NUMBER OF PRINCIPAL COMPONENTS AND VARIANCE EXPLAINED

		Vibration	Current	Micro-flown	Micro-phone	Voltage
Test A	PC	30	18	20	30	28
	σ expl.	92%	95%	90%	95%	91%
Test B	PC	22	12	14	25	15
	σ expl.	89%	91%	86%	94%	82%
Test C	PC	23	16	14	27	29
	σ expl.	88%	94%	84%	94%	92%

This test case investigates the performance of the algorithm for loading conditions exceeding those considered in the training set and for unique environmental conditions, specifically when the background noise is at increased levels. The results are shown in Table III. The correct diagnosis of the healthy motor was 100%, as well as the diagnosis for F_1, F_4, F_5, F_6, and F_7. In case of stator fault F_2, there is a 2% missed alarm rate. In case of stator fault F_3, the algorithm misdiagnoses F_3 as F_4 in 8% of the cases. These phenomena are similar to those observed in Test Case A: the stator faults are less severe and less easy to diagnose. Due to fault similarities, the algorithm can sometimes misdiagnose stator fault severities or confuse them with the healthy motor. The F_1 score is 99.88%, which is even higher than the random split test case.

D. Principal Components

The number of principal components is shown in Table IV for each signal type together with the variance explained to complement the results in the above-presented test cases. In case of the random split in Test Case A, the variance explained by the chosen principal components is always above 90%. In case of Test Case B and C, the number of chosen principal components is less than for Test Case A. This is due to the specific loading and noise conditions chosen for the test sets.

The first few principal components have been analyzed for all signal types to determine if there is any feature that dominates the principal component coefficients in the loading matrix. It was found that there was no single feature that would stand out for any signal type, therefore the importance of PCA for correlation reduction is further confirmed.

Fig. 4 shows the first principal components of the five signal types, for all observations through the 58 datasets. The principal component values were obtained from the normalized feature values as described in Section II-D. In comparison to Fig. 3,

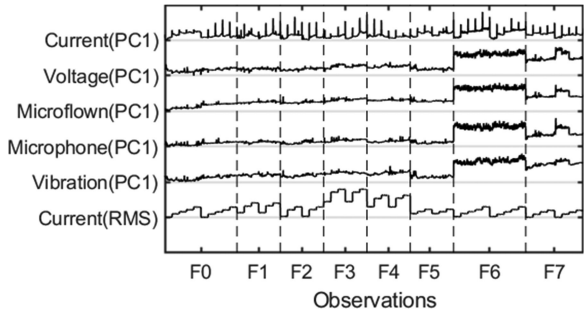

Fig. 4. First principal components of the 5 different signal types, for all observations through the 58 datasets, the rms of the current is given as reference for the loading conditions.

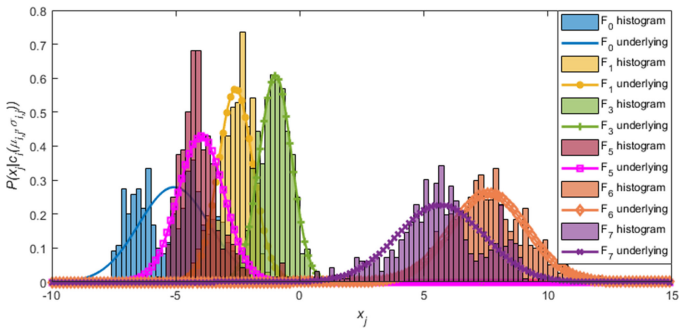

Fig. 5. Histograms and underlying normal distributions of the first principal component of the vibration signal by fault conditions.

where the rms of the five signal types are shown, it may be observed that the load dependence of the signals is less evident in the principal components. This further justifies the application of PCA for problems where the analyzed problem contains data from several loading conditions.

Fig. 5 presents the histograms and underlying Gaussian distributions of the first principal component of the vibration signal by fault conditions. The distributions for each fault types have distinct mean and variance values and are not significantly different from Gaussian distributions. It can be observed that F_6 and F_7 are the most distinguishable from F_0, whereas the other stator faults have overlaps with F_0. It should be noted that F_0 shows the evidence of multimodal behavior. This is due to the additional background noise incorporated to investigate the influence of different environmental conditions on the accuracy of diagnosis. However, as shown in Sections V-A–V-C, this noise did not significantly influence the resulting likelihood calculations.

E. Single-Stage Data Fusion

A comparison of the performance of the two-stage approach relative to a more standard single-stage approach, where sensors are not separated according to type, but instead all fused in a single stage, was performed. The total 3480 observations were randomly split according to the conventional 70–30% partition to training set and test set. The random split was applied 100 times to a single-stage approach and the averaged results are shown in Table V. The results show that the performance of the single-stage algorithm significantly drops compared to the

TABLE V
SINGLE-STAGE DATA FUSION

		Diagnosed condition							
		F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7
Actual condition	F_0	0.09	0.10	0.32	0.05	0.11	0.32	0.01	0.00
	F_1	0.02	0.94	0.01	0.01	0.00	0.01	0.01	0.00
	F_2	0.03	0.01	0.90	0.02	0.01	0.03	0.00	0.00
	F_3	0.02	0.00	0.00	0.93	0.05	0.00	0.00	0.00
	F_4	0.02	0.00	0.01	0.16	0.77	0.03	0.01	0.00
	F_5	0.01	0.01	0.01	0.01	0.01	0.94	0.01	0.00
	F_6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	F_7	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.98

TABLE VI
PROPORTION OF CORRECT DIAGNOSES FOR EACH FAULT TYPE WHEN CONSIDERING EACH SIGNAL INDIVIDUALLY AND AFTER TWO-STAGE FUSION

	Single-Stage Fusion by Signal Type					Two-Stage fusion
	Vibration	Current	Micro-flown	Micro-phone	Voltage	
F0	0.75	0.88	0.91	0.73	0.66	0.94
F1	0.92	0.98	0.90	0.90	0.83	1.00
F2	0.87	0.62	0.84	0.82	0.85	0.98
F3	0.86	0.82	0.91	0.90	0.86	0.98
F4	0.87	0.72	0.87	0.88	0.88	0.98
F5	0.90	0.92	0.94	0.92	0.90	1.00
F6	1.00	0.89	0.98	0.97	0.99	1.00
F7	1.00	0.96	0.91	0.99	1.00	1.00

results of the two-stage method shown in Table I. The most significant difference appears in the reduced successful detection of the healthy motor, with the single-stage approach yielding false alarms in 91% of test cases. The F_1 score is 92%.

F. Comparison of Results With SVM

To provide a quantitative comparison with another classifier, the proposed PCA and two-stage Bayesian method is compared with the well-known SVM. Test Case A, B, and C are repeated using the default *fitcecoc* MATLAB implementation of the SVM for multiclass problems with one against one classification strategy and a linear kernel function. The F_1 scores are compared. Similarly to the investigation described in Section V-F, the SVM was applied in a single stage. A 70-30% data split was applied and repeated 100 times resulting in a 99.96% F_1 score for Test Case A. This result is 0.64% better than that of the proposed method. For Test Case B, the F_1 score for the SVM was 96.15%, which is 1.84% below than what was achieved with the newly proposed method. For Test Case C, the F_1 score for the SVM was 97.8%, which is 2.08% below than what was achieved with the newly proposed method. While the performance of the two approaches is comparable, an advantage of PCA and two-stage Bayesian method lies in its transparency and modularity. Furthermore, the method also provided a marginally improved performance in the case of environmental and loading conditions not contained in the training set, as shown in Test Cases B and C.

G. Signal Types Separately Versus Two-Stage Fusion

Table VI shows the performance of only considering a single-stage fusion of features from a single-signal type, for the random split Test Case A. For comparison, the equivalent performance from the two-stage approach, which fuses the data from all sensors types in the global fusion stage, is also given. Results are given in terms of proportion of correct diagnoses, which are equivalent to the values on the diagonal of the previously presented results (see Tables I–III). It is evident that the two-stage data fusion of multiple signal types outperforms the equivalent results when only considering a single-signal type. This is due to the fact that the different sensor types have different strengths and weaknesses. For example, it may be observed that the analysis based only on vibration signals accurately diagnosed the mechanical bearing fault F_7 in 100% of test cases, but was only able to diagnose an electrical stator fault, such as F_1, in 92% of cases. In contrast, when only current signals were considered, stator fault F_1 was diagnosed correctly in 98% of cases, but bearing fault F_7 was only diagnosed correctly in 96% of cases. When the two signals are fused, the conditional probabilities in the global confusion matrix effectively gives greater weight to vibration signals and less weight to current signals when diagnosing mechanical faults and vice versa in the case of diagnosing electrical faults. This leverages the strengths of each sensor type for fault monitoring and minimizes the impact of the weaknesses.

VI. DISCUSSION

In this section, the results and the structure of the algorithm are discussed further, highlighting the observed strengths and weaknesses of the algorithm.

A. Implementation and Constraints

The training of the method takes place offline using historical datasets containing healthy and faulty data. Once the model is trained, diagnosis can be performed either online or offline. By applying a sliding window of the same size as used for training, the new sensor measurements can be fed into the two-stage Bayesian classifier online after the feature extraction and PCA steps have been performed. The width of the window could be different based on the nature of the monitored system, the extracted features, and the data available. The computational complexity of the classifier is proportional to the number of principal components retained and the number of fault modes monitored. The computational complexity of the feature extraction and PCA step depends on the number of features extracted and the size of the sliding window. For a better representation of the original feature space, nonlinear multivariate methods, such as kernel PCA [33], could be explored in the future instead of the currently used linear PCA. While it falls out of the scope of this paper, it should also be noted that the features used as inputs to the method may also be refined according to state of the art signal processing and feature extraction methods so that they may better discriminate between different health states. Thus, the accuracy and reliability of the approach would likely be improved further. In (4) and (8), the likelihoods might result in very small values if the number of features m, the number of sensors N, or the number of fault cases M is large. To avoid numerical problems, a logarithmic formulation might be considered.

B. Algorithm Validation

In Section V, three different algorithm validation test cases were presented by splitting the data into different training sets, test sets, and validation sets. It has been shown that for small datasets, the simple split-sample estimates can be biased and cross validation is more suitable for the prediction assessment of the classifiers [34]. In the case of a two-stage method, cross validation is unfeasible due to the increase in the number of computational steps associated with the addition of the global fusion stage and the use of a validation set. Specifically, relative to a simple single-stage fusion, when implementing cross validation on a two-stage approach, the method becomes n^2 more computationally expensive, where n is the number of the observations, as both the local and the global stages have to be trained using separate training sets. In this paper, a pragmatic split-sample method was considered. It is also foreseen that such an approach would be applicable for applications of the method with larger volumes of datasets available. In the future, increases in computing power might also allow the cross-validation approach to be feasibly applied.

C. Naïve Bayes Classifier Using Kernel Density Estimate (KDE)

The GNB classifier is a parametric method that assumes a normal distribution of the observation variables. The more the distribution of the observation variables differs from the normal distribution, the less accurate the method is. One possible way to eliminate this Gaussian assumption is to use a naïve Bayes classifier with KDE, where the probability density function of the features are estimated using a nonparametric kernel distribution. Such an approach can be used when there is no prior knowledge regarding the distribution of the data, no assumptions are made, or a parametric distribution cannot describe the data. Tests conducted using such a naïve Bayes classifier with KDE, with the same random split as described in *Test Case A*, yielded comparable results to the GNB classifier. The naïve Bayes classifier with KDE resulted in correct classification rates in the $\pm 2\%$ range compared to the results in Table I, whereas the F_1 score is 99.64%, which is 0.32% better compared to the results in Table I. However, when applying KDE, the computation time was two magnitudes greater for the local stage than for the case of the GNB classifier. It took 4.277 s for the original method to train the local stage and obtain the confusion matrixes for the vibration signals, whereas the same computation took 351.78 s with KDE. The processing hardware was an Intel Core i5-4300U, 1.9 GHz.

D. Two-Stage Data Fusion Without PCA

While not the primary focus of this paper, it is worth noting that an investigation into the importance of incorporating the PCA step into the algorithm was also performed. It was observed that when the PCA step was omitted from the algorithm, all test cases, including fault cases, were subsequently diagnosed as being healthy (F_0). This was due to the load dependence of the features. This observation indicates that a PCA step, or similar, ensures that the algorithm is robust against changing loading and environmental conditions.

E. Advantages of the Method

The preceding sections provide quantifiable comparisons of the performance of the algorithm when including the novel steps of applying a GNB classifier and splitting the approach into two stages, relative to the cases when the steps are omitted. Due to the multitude of ways of properly designing and tuning various algorithms, it is unfeasible to perform similarly rigorous quantitative comparisons to benchmark the method relative to other data-driven fault detection methods. However, qualitative comparisons, which can guide design decisions at an early stage of the analytics development process, can be made. The main advantages of the proposed method are its transparency and modularity. In contrast to many other data-driven fault diagnosis methods, such as SVMs or neural networks, the decision-making process of the algorithm is easily back traceable from the global predictions to the inputs of the local stage to identify how the different sensors reacted to a fault. Such transparency is important for cases where the algorithm will be used to support maintenance decisions. While in this paper, only MAP probabilities were considered, in practice, the Bayesian sensor fusion approach allows the results to be presented in the form of likelihoods, showing the probability of each fault condition being present. Again, this additional insight can support maintenance decisions.

The modularity of the approach, achieved by splitting the data fusion into two stages, also offers further advantages when considering practical implementation. In the case of a sensor being removed from a system, there is no need to retrain the whole model, as the removed sensor type can easily be omitted from the decision-level fusion. This is not possible for other fault diagnosis methods that only consider feature-level data fusion. Similarly, additional sensor types may be readily incorporated into the analysis with limited requirements for retraining. Recently, a trend of monitoring the health of components via signals recorded from connected elements, for example, monitoring gearboxes and bearings via electrical signals recorded from connected electrical motors, has emerged [35], [36]. Such emerging methods could also easily be incorporated into the algorithm, serving as an additional source of information for further improving the accuracy of diagnosis.

VII. Conclusion

In this paper, the performance of a newly proposed PCA and two-stage Bayesian sensor fusion method was evaluated under various test scenarios. The algorithm was shown to be able to diagnose stator faults, broken rotor bar faults, and bearing faults in induction motors, with low false and missed alarm rates. The algorithm also proved its ability to diagnose faults under different loading and environmental conditions. In addition to discussing the several advantages of the presented method, the limitations of the method were also highlighted. For example, it was shown that the method is capable of correctly distinguishing different types of fault, however, to consistently distinguish between

different fault severities, adequate training sets are required at comparable loading conditions.

In the future, the algorithm can potentially be extended so that it may be used not only with steady-state signals. Additionally, the performance of the method may be refined by further tailoring the extracted features to the monitored system. It was shown that by fusing data recorded from different sensor types, the proposed method is capable of diagnosing both mechanical and electrical faults. In the future, the algorithm should also be tested for other fault detection and condition-monitoring scenarios, for example, in process-monitoring applications.

Acknowledgment

The authors would like to thank M. Sułowicz, K. Weinreb, J. Petryna, and A. Dziechciarz, from Cracow University of Technology, and W. Batko, M. Kłaczyński, J. Wierzbicki, T. Wszołek, and J. Frączek, from AGH University of Science and Technology, for carrying out the measurement campaign.

References

[1] S. Karmakar, S. Chattopadhyay, M. Mitra, and S. Sengupta, *Induction Motor Fault Diagnosis: Approach Through Current Signature Analysis*. New York, NY, USA: Springer, 2016.

[2] P. F. Albrecht, J. C. Appiarius, R. M. McCoy, E. L. Owen, and D. K. Sharma, "Assessment of the reliability of motors in utility applications - Updated," *IEEE Trans. Energy Convers.*, vol. EC-1, no. 1, pp. 39–46, Mar. 1986.

[3] Motor Reliability Working Group of IEEE, "Report of large motor reliability survey of industrial and commercial installations," *IEEE Trans. Ind. Appl.*, vol. IA–4, no. 4, pp. 853–872, Jul./Aug. 1985.

[4] *IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems*, IEEE Std 493–1997, 1997.

[5] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.

[6] W. Li, "Detection of induction motor faults: A comparison of stator current, vibration and acoustic methods," *J. Vib. Control*, vol. 12, no. 2, pp. 165–188, 2006.

[7] M. R. Mehrjou, N. Mariun, M. Hamiruce Marhaban, and N. Misron, "Rotor fault condition monitoring techniques for squirrel-cage induction machine—A review," *Mech. Syst. Signal Process.*, vol. 25, no. 8, pp. 2827–2848, 2011.

[8] A. Siddique, G. S. Yadava, and B. Singh, "A review of stator fault monitoring techniques of induction motors," *IEEE Trans. Energy Convers.*, vol. 20, no. 1, pp. 106–114, Mar. 2005.

[9] A. Glowacz, "Fault diagnosis of single-phase induction motor based on acoustic signals," *Mech. Syst. Signal Process.*, vol. 117, pp. 65–80, 2019.

[10] A. Glowacz, "Acoustic based fault diagnosis of three-phase induction motor," *Appl. Acoust.*, vol. 137, no. 11, pp. 82–89, 2018.

[11] B. S. Yang and K. J. Kim, "Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 403–420, 2006.

[12] M. S. Safizadeh and S. K. Latifi, "Using multisensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Inf. Fusion*, vol. 18, no. 1, pp. 1–8, 2014.

[13] M. Khazaee, H. Ahmadi, M. Omid, A. Moosavian, and M. Khazaee, "Classifier fusion of vibration and acoustic signals for fault diagnosis and classification of planetary gears based on Dempster–Shafer evidence theory," *Proc. Inst. Mech. Eng. E, J. Process Mech. Eng.*, vol. 228, no. 1, pp. 21–32, 2014.

[14] T. H. Loutas, D. Roulias, E. Pauly, and V. Kostopoulos, "The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery," *Mech. Syst. Signal Process.*, vol. 25, no. 4, pp. 1339–1352, 2011.

[15] D. L. Hall and J. Llinas, "An Introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[16] R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP Ann., Manuf. Technol.*, vol. 59, no. 2, pp. 717–739, 2010.

[17] T. Sutharssan, S. Stoyanov, C. Bailey, and C. Yin, "Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms," *J. Eng.*, no. 7, pp. 215–222, 2015.

[18] B.-S. Yang, T. Han, and Z.-J. Yin, "Fault diagnosis system of induction motors using feature extraction, feature selection and classification algorithm," *JSME Int. J. Ser. C*, vol. 49, no. 3, pp. 734–741, 2006.

[19] M. Farajzadeh-Zanjani, R. Razavi-Far, and M. Saif, "Dimensionality reduction-based diagnosis of bearing defects in induction motors," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2017, pp. 2539–2544.

[20] J. J. Saucedo-Dorantes, M. Delgado-Prieto, R. A. Osornio-Rios, and R. De Jesus Romero-Troncoso, "Multifault diagnosis method applied to an electric machine based on high-dimensional feature reduction," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3086–3097, May/Jun. 2017.

[21] P. Gangsar and R. Tiwari, "Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms," *Mech. Syst. Signal Process.*, vol. 94, pp. 464–481, 2017.

[22] R. H. C. Palacios, I. N. Da Silva, A. Goedtel, W. F. Godoy, and T. D. Lopes, "Diagnosis of stator faults severity in induction motors using two intelligent approaches," *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 1681–1691, Aug. 2017.

[23] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 724–739, 2009.

[24] K. Tidriri, N. Chatti, S. Verron, and T. Tiplica, "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges," *Annu. Rev. Control*, vol. 42, pp. 63–81, 2016.

[25] V. H. Jaramillo, J. R. Ottewill, R. Dudek, D. Lepiarczyk, and P. Pawlik, "Condition monitoring of distributed systems using two-stage Bayesian inference data fusion," *Mech. Syst. Signal Process.*, vol. 87, pp. 91–110, 2017.

[26] A. Stief, J. R. Ottewill, M. Orkisz, and J. Baranowski, "Two stage data fusion of acoustic, electric and vibration signals for diagnosing faults in induction motors," *Elektron. ir Elektrotechnika*, vol. 23, no. 6, pp. 19–24, 2017.

[27] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2010.

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: *Wiley*, 1995, pp. 51–65.

[29] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[30] N. Friedman, D. Geiger, and M. Goldszmit, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, pp. 131–163, 1997.

[31] K. Weinreb, M. Sułowicz, and J. Petryna, "Faults detection in cage induction motor with parallel branches," *Tech. Trans., Czas. Tech.*, vol. 2-E, pp. 53–64, Dec. 2016.

[32] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[33] S. W. Choi, C. Lee, J. M. Lee, J. H. Park, and I. B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics Intell. Lab. Syst.*, vol. 75, no. 1, pp. 55–67, 2005.

[34] A. M. Molinaro, R. Simon, and R. M. Pfeiffer, "Prediction error estimation: A comparison of resampling methods," *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005.

[35] S. H. Kia, H. Henao, and G. A. Capolino, "Trends in gear fault detection using electrical signature analysis in induction machine-based systems," in *Proc. IEEE Workshop Elect. Mach. Des., Control, Diagnosis*, 2015, pp. 297–303.

[36] J. R. Ottewill and M. Orkisz, "Condition monitoring of gearboxes using synchronously averaged electric motor signals," *Mech. Syst. Signal Process.*, vol. 38, no. 2, pp. 482–498, 2013.

Anna Stief was born in Debrecen, Hungary. She received the B.S. and M.S. degrees in electrical engineering from the Budapest University of Technology and Economics, Budapest, Hungary, in 2013 and 2015, respectively. She is currently working toward the Ph.D. degree in automatics control and robotics with the AGH University of Science and Technology, Kraków, Poland.

She is currently a Scientist with ABB Corporate Research Center, Kraków, Poland. Her research interests include heterogeneous data fusion, machine learning, and advanced analytics applied to condition-monitoring applications.

James R. Ottewill (M'18) was born in London, U.K. He received the B.Eng. (Hons.) and Ph.D. degrees in mechanical engineering from the University of Bristol, Bristol, U.K., in 2005 and 2009, respectively.

He is currently a Senior Principal Scientist with ABB Corporate Research Center, Kraków, Poland, working in the field of applied analytics for condition-monitoring applications. His main research interests relate to advanced approaches for diagnostics and prognostics including modeling and analysis of nonlinear systems, signal processing, and information fusion.

Jerzy Baranowski (M'11) was born in Kraków, Poland. He received the master's, D.Phil., and D.Sc. degrees in automatics control and robotics from the AGH University of Science and Technology, Kraków, Poland, in 2006, 2010, and 2017, respectively.

He is currently an Assistant Professor with the Department of Automatic Control and Robotics, AGH University of Science and Technology. His current research interests include fault detection, Bayesian statistics, and applications of fractional calculus.

Dr. Baranowski is an Academic Editor for the *Mathematical Problems in Engineering*. He was the recipient of the scholarship of Polish Ministry of Science and Higher Education for outstanding young scientists in 2017.

Michal Orkisz (SM'16) was born in Kraków, Poland. He received the B.S. degree in computer science, physics, and mathematics and the Ph.D. degree in condensed matter physics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1988 and 1994, respectively.

He was with Gel Sciences, Inc., and Genome Therapeutics, Bedford, MA, USA. Since 1998, he has been with ABB's Corporate Research, Krakow, Poland, where he holds a Senior Principal Scientist position and has dealt with tools development and analysis for condition monitoring, diagnostics, risk analysis, as well as computer vision.