Data Analytics Towards Predictive Maintenance for Industrial Ovens

A Case Study Based on Data Analysis of Various Sensors Data

Vaia Rousopoulou[1(✉)], Alexandros Nizamis[1], Luigi Giugliano[2], Peter Haigh[3], Luis Martins[4], Dimosthenis Ioannidis[1], and Dimitrios Tzovaras[1]

[1] Centre for Research and Technology Hellas-Information Technologies Institute (CERTH/ITI), 57001 Thessaloniki, Greece
`vrousop@iti.gr`
[2] MLW, Links Foundation, Turin, Italy
[3] Tyndall National Institute, Lee Maltings, Cork T12 R5CP, Ireland
[4] Boston Scientific Limited, Clonmel E91 T862, Ireland

Abstract. In Industry 4.0, predictive maintenance aims to improve both production and maintenance efficiency. The interconnected machines and IoT devices produce a variety of data that enable the early detection of anomalies and failures by predictive analytic algorithms. Predictive analytics can also reduce the machines downtimes and decrease the production of faulty products. This paper introduces predictive analytics for industrial ovens and their application in a real-world's oven used by a leading medical devices manufacturer. Two distinct approaches are presented in this work. A technique based on existing sensors for oven failure prediction based on monitoring and log data; and a technique based on deployed sensors for fault diagnosis based on acoustic data. Deep learning techniques have been applied on existing sensor and event log data, especially temperature monitoring, whereas an outlier detection analysis were implemented on acoustic sensor measurements. Both analytics methods create a complete solution able to detect early oven failures from their root.

Keywords: Industry 4.0 · Predictive maintenance · Deep learning · Outlier detection · Acoustic sensors

1 Introduction

In Industry 4.0, the interconnectivity of machines and systems, the installation of IoT devices and the development of novel predictive algorithms enable predictive maintenance processes. In particular, in the case of printed circuit boards' (PCB) production, data from built-in sensors from ovens can be used alongside with data from externally installed sensors. These data sources enable the creation of a variety of big datasets that can be used by data analytics tools in order to boost predictive maintenance processes and enhance the decision-making in the shop-floor.

In this case study, the predictive maintenance procedures are focused on PCBs production line and the ovens that were used on Boston Scientific facilities in Clonmel (BSL). Boston Scientific is one of the largest medical device companies in the world. It manufactures implantable electronics that use printed circuit boards (PCB) to connect and physically support all the necessary electronic components, in which solder paste is used to temporarily attach these components to the electric contact pads onto the PCB into a final circuit assembly (PCBA). Once that assembly is complete, and in order to make this attachment permanent, this assembly is then passed through a reflow oven[1], an equipment that uses heat to transform the solder paste into a permanent joint and obtain a PCBA ready for the next steps of manufacture. To achieve an even heat distribution as well as maintain the interior temperature bellow a safety threshold, reflow ovens also include several blower fans on the inside.

In the manufacturing environment of BSL, it is normal and expected for small malfunctions with the blower fans to occur from time to time. However, their continuous operation at suboptimal conditions gradually leads to instability inside the oven. If this fact remains unnoticed, the machine deteriorates over time, thus leading to failures. Higher than normal or uneven temperatures can cause malfunctions and breakdowns on the machines. Also, a malfunction inside the oven leads to poorly built PCBAs that will then be classified as non-compliant (NC) and considered scrap to be thrown away. As the products built at BSL have a very high cost, it is extremely important to prevent as many NC's as possible, as well as to understand what causes them. This helps resolving systemic faults along the process steps which can lead to huge savings in manufacturing cost. For every actual failure (motor/blower fan failure) the company loses thousands of euros and some millions of them in yearly basis.

In the reflow ovens, several parameters like temperature, pressure and power consumption are constantly being monitored and logged. However, due to their daily experience with the equipment, the operators are able to realize, even before the NC's start to appear, when a reflow oven is close to having a malfunction due to fan failure. When these fans begin to fail, their motors often start making high pitched noises. Based on this, acoustic data was therefore defined as useful data to collect and use in combination with the data the oven itself already outputs.

Prediction of the blower fan failures in these ovens is therefore extremely important to BSL both from quality and cost perspectives. In order to achieve this, a combination of the data outputted from those reflow ovens with additional acoustic data outputted from installed acoustic sensors will be used by a variety of analytic tools. These tools will apply different algorithms and methodologies for prediction of failures and detection of abnormalities and outliers on real time. The analytics output would be used by decision support systems that offer monitoring, display abnormalities and generally provide a status update to the maintenance manager in a near real-time manner. The analytics output is also

---

[1] http://www.hellerindustries.com/pdf/1809EXL_SPEC_SHEET.pdf.

possible to support the maintenance team's decision about when and what to replace before failure occurs, helping them to determine the optimal time of replacement for each blower fan.

In this paper, a complete solution for predictive maintenance for the aforementioned ovens based on the analysis of acoustic sensors data and data from ovens' built-in sensors is presented. By providing the analysis outcomes to a company's decision support system, time and costs are saved in maintenance procedures since failures can be predicted and the maintenance plan adjusted accordingly, thus boosting an overall equipment efficiency, preventing unnecessary NC product and decreasing the overall scrap costs.

The paper is structured as follows. Following the Introduction, a related review is presented. The deployed acoustic sensors and the analysis of the fans acoustic data are presented on Sect. 3. Section 4 contains a thorough analysis of ovens' events based on deep learning techniques. Finally, the conclusions and are drawn at Sect. 5.

2   Related Work

While various approaches have been put forward to address the issue of predictive maintenance through fault diagnosis in industrial machines, in the following we will focus on a selection of literature that investigate condition based maintenance depended on log event monitoring in addition to non-invasive maintenance techniques. The work of [1] is a review of developments in condition based maintenance field, emphasizing on three key steps: data acquisition, data preprocessing and decision-making step. The writers expound on the combination of analyzing event data and condition monitoring data together in time and frequency domain and point out the advantages of multi-sensorial data fusion for accurate condition monitoring, fault diagnosis and prognosis. In the same regard, the work of [2] presents a data-driven approach based on multiple-instance learning for predicting equipment failures by mining equipment event logs. The proposed method of [2] utilizes state-of-the-art machine learning techniques to build predictive models from log data.

Additionally, a comprehensive state-of-the-art analysis of predictive maintenance techniques is conducted in the work of [3], indicating the need of dealing with random failures of industrial equipment. This research defines three major types of predictive maintenance technologies, based on their data sources: the existing sensor-based maintenance technique, the test-sensor-based maintenance technique and the test signal based maintenance technique. The first two categories that concern our research consist of methods that use data from existing process sensors that measure variables like temperature, pressure, level, and flow and methods that use data from test sensors such as accelerometers for measuring vibration and acoustic sensors for detecting leaks, respectively.

A pertinent work, carried out recently is [4], presenting an acoustic sensing system for industrial fans fault detection. The writers highlight the ability of the acoustic sensors to give a better high-frequency response to the system.

The system uses the combination of fan rotation speed and its corresponding sound and conduct filtering and abnormal feature analysis, resulting in an effective, real-time early fault detection and prediction system.

Last but not least, a systematic study on fault diagnosis was carried out by Glowacz [5–7]. His preliminary work in the field is [5], where fault diagnosis of electric motors through acoustic signals is presented, implementing feature extraction from acoustic signals of faulty and faultless motors and pattern recognition methods to detect imminent flaws. More recent works are [6] and [7], which focus on fault diagnostic techniques based on acoustic signals for induction motors. In both works, the method pipeline consists of the three signal processing steps: pre-processing, feature extraction, classification. The acoustic signal is recorded, transformed into smaller audio files and processed by implementing amplitude normalization and Fast Fourier transform. Then characteristic features are extracted from the acoustic signals and classification is performed with Nearest Neighbour and other classifiers, resulting in a system capable of detecting unexpected failures.

3   Acoustic Data Analysis Techniques

In order to take the prediction level in an earlier phase, we deployed acoustic sensors in the ovens. We selected the use of acoustic test sensors intrigued by the advantages of them: the easy accommodation, the low-cost equipment and the accessible data information that they provide.
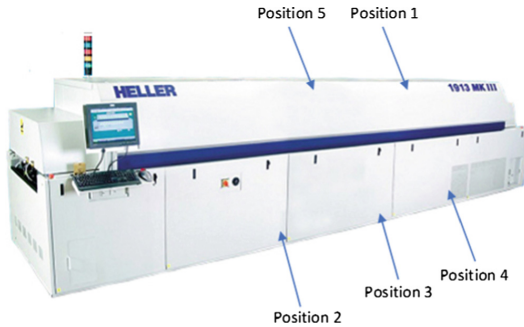

Fig. 1. Positions of deployed acoustic sensors on the oven.

3.1   Acoustic Sensors Deployment

Typical solder reflow ovens of the type deployed at Boston Scientific have over 30 fans within them. The key objective of the condition monitoring system is;

– To alert the operating technician that there is a high potential of failure about to occur.

– To notify the maintenance technician which area (zone) within the oven the faulty fan is located.

The acoustic sensing deployment consists of five sensors placed within specific zones that is within hearing distance of all 30 fans as shown in Fig. 1. The zonal approach to sensing failures also reduces system cost and complexity.

The design approach of the sensing system was to use standard equipment as well as open standards based protocols and formats. This enables greater flexibility in the tools to process the data and provides for ease of use to future development of the system. Closed bespoke systems can be difficult to maintain as well as harder to adopt more broadly outside the initial deployment.
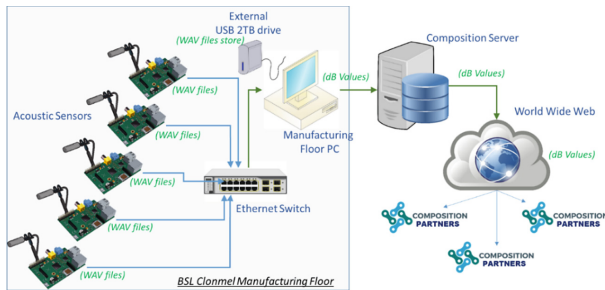

Fig. 2. Hardware architecture overview.

Figure 2 gives an overview of the hardware architecture. The acoustic sensors are constructed with a Raspberry Pi integrated with a SPH0645LM4H microphone via an I2C interface. The sensors record 20 s of sound every 5 min in a standard PCM format. This data is sent in an uncompressed WAV format to the Manufacturing floor PC via Ethernet. The WAV files are processed on the PC where each 20 s recording is converted into a single amplitude value. The WAV file is stored locally on a USB drive and the single amplitude value is stored on a server remote from the factory.

In this work, the algorithms require a single amplitude value in dBSPL. This value is calculated by taking the mean of the peak envelope filtered data and then converting to a dBSPL value. Raspberry Pi sensors record 20 s of sound every 5 min via a shell-script that runs in a continuous loop. The data is timestamped and saved locally in a WAV file format. Thereafter, a local PC is used where a continuous running script collects the newest WAV files from each of the five Raspberry Pi, and then a MATLAB script that passes the data through a peak envelope filter and then takes the mean value to produce a single dB amplitude value. For each set of 5 measurements a single CSV file is generated that stores the 5 values from each sensor. The timestamp in the filename is identical to that of the WAV files and so cross correlation is possible at a later date if required. Then, this data is copied to the server in order to be accessible from analytics tools.

3.2   Acoustic Outlier Analysis for Failure Detection

A method for predictive maintenance in industrial ovens is presented hereby. The proposed method is based on outlier detection out of acoustic measurements. The dB amplitude values derived from the deployed acoustic sensors, as described in Sect. 3.1, formulate the required dataset. This dataset suffers from a pitfall: the industrial oven cannot fail intentionally for the purpose of the specific research, thus the acoustic measurements entail only healthy information. In order to address the challenge of absence of faulty acoustic measurements, we propose an outlier analysis of the imbalanced dataset followed by implementation of well-known classification techniques. The scope of the proposed method is to detect observations in the audio data measurements which deviate so much from the other samples as to indicate that there might exist a possible failure in the ovens. An operational period can be considered as an outlier and indicate an upcoming abnormality. The calculated outliers will constitute the false class of the upcoming classification and prediction process.

Outlier Detection. Three outlier detection algorithms where implemented and compared in this respect: Mean Absolute Deviation (MAD), Local Outlier Factor (LOF) [8] and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [9]. The MAD value of a signal is calculated over a rolling window with a fixed number of data points of the sample. The rejection criterion of the MAD value is defined based on [10], meaning that the values that lie between specific limits are considered outliers. Next, LOF algorithm is implemented to compute the local density deviation of an acoustic data sample with respect to its neighbors. It considers as outliers the samples that have a substantially lower density than their neighbors. The last outlier detection algorithm used is DBSCAN. The data are fitted to DBSCAN, clusters are created and each acoustic sample is assigned to a cluster. The number of clusters is estimated automatically and outliers (noise) are assigned to the -1 cluster. The parameters of DBSCAN are selected after a few trials.
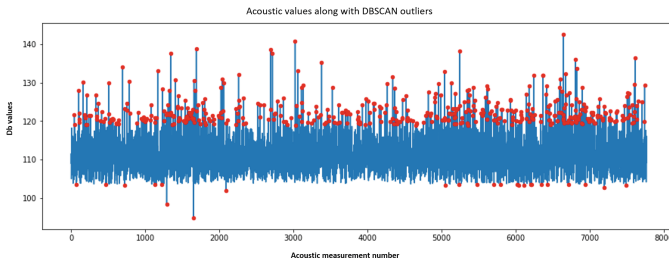

Fig. 3. The outliers detected by DBSCAN.

The three algorithms search for samples that deviate from the majority. The lack of faulty data lead to a quite small number of detected outliers, though DBSCAN identified the most outlier points, 1149 out of 7767 samples, which are noted as the faulty class for the classification process (Fig. 3).

SVM Classification. The previously described algorithms are used for indicating noisy points in the acoustic samples. The outliers are characterized as faulty data in order to create a binary class problem. However, the faulty data are only 14% of the overall data points, forming an imbalanced dataset. A classifier fed with this dataset will be more sensitive to identify the majority class of non-faulted fans and the classification will be biased always predicting the positive class. In order to address the issue of imbalanced dataset, an oversample of the minority class is performed. Synthetic Minority Oversampling Technique (SMOTE) is used to resample and synthesize new elements for the minority class, based on those that already exist [11]. A minority class point is chosen randomly and the k-neighbours of it are calculated. The synthetic data are placed between the picked point and the calculated neighbours. SMOTE technique diminishes the issue of overfitting as the new synthetic data are created rather than reproducing samples and corrupting information. The oversampling of the detected outliers (red points) is illustrated below.
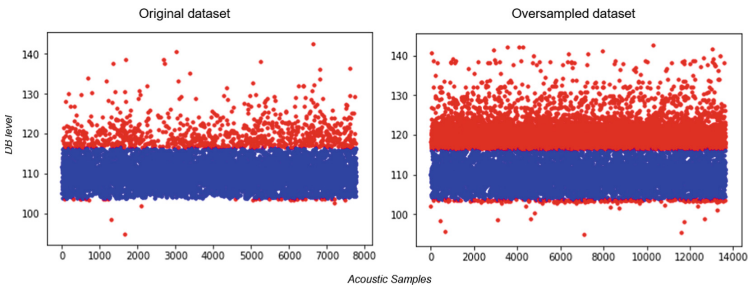

Fig. 4. The dataset oversampling with SMOTE.

The original dataset in shown in the left sub-figure of Fig. 4, where the positive class consists of the 86% of the dataset and the negative class is the 14% respectively. After implementing SMOTE for oversampling the two classes contain 50% of the samples each (right sub-figure of Fig. 4). An SVM classifier is used for training the new dataset. The two possible label classes are faulted acoustic sample (=1) and non-faulted acoustic sample (=0). A training dataset is created including the 70% of the overall dataset and the rest 30% is used for the testing dataset. The SVM model is trained with the following parameters: cost = 0.5 and kernel = Radial basis function kernel (RBF).

In order to evaluate the trained model the testing dataset is used to predict the class of the points based on their decibel values. Figure 5 is the illustration

of the confusion matrix, to evaluate the quality of the output of the SVM classifier on the acoustic data set. The diagonal elements represent the number of points for which the predicted label is equal to the true label, while off-diagonal elements are those that are mislabelled by the classifier. The diagonal values of the confusion matrix are quite high, indicating many correct predictions of the faulted and not-faulted acoustic data.
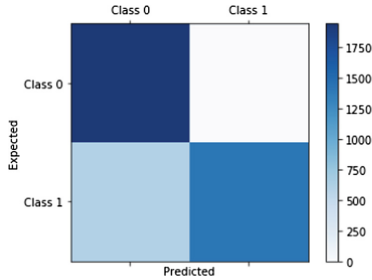

Fig. 5. The confusion matrix of the SVM classifier on the acoustic dataset.

The evaluation metrics are accuracy, precision, recall and F1 score. Accuracy is the ratio of correctly predicted observation to the total observations. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. Recall is the ratio of correctly predicted positive observations to the all observations in actual positive class. The F1 score is a weighted average of the precision and recall.

Table 1. SVM model evaluation metrics.

Metrics	Result
Accuracy	0.85
Precision	0.76
Recall	1.0
F1-score	0.86

The evaluation metrics were calculated and the results indicate that the model is capable of detecting faulted measurements (Table 1). All of the metrics are over 75%. Note that the recall metric reaches 1.0 because of the fact that the non-faulted data are generated and do not exist in the initial dataset. New measurements can be given to the SVM prediction model, optimally live data, that can predict if there exist a possible failure in the reflow ovens with 85% accuracy.

SVM Prediction Model. For each audio sensor, an SVM prediction model is extracted according to the aforementioned classification analysis. The purpose of the proposed method is to be applied in live audio measurements and detect possible failures in reflow ovens. The SVM model of each sensor receives as input the result value of the audio measurement, namely a timestamped decibel level. The model decides whether the decibel measurement belongs to the faulted or non-faulted class. In case that the model detects the faulty class for five consecutive measurements, the system informs that there is a possible abnormal function in the oven.

4   Existing Oven Sensor and Log Data Analysis Techniques

This section describes a deep learning method for predictive maintenance of industrial ovens based on log data and events. The log data are collected as part of the operational routine for monitoring the system performance. The proposed method indicates that the value of these data can be exploited to serve as indicators of a system' s health condition.

Data Acquisition. BSL provided sensor data files containing the temperature set by the user, the measured temperature and the output power at the solid state relay of the reflow, and event files containing a list of failure events related to the oven. The dataset has been created by merging the data files and the event files using the timestamp value as key. The event with the closest timestamp is assigned to each data entry in the data file, duplicating the available events. The process follows the creation of each single entry and the matching of events for each data entry. The selection of this approach was based on two facts: (a) there are more data entries than event entries for each file; there is the assumption that if an event occurs at the time t, (b) there is a high probability that the same event "was occurring" or about to occur at the time $t - 1$ or $t + 1$ since the sampling time is short.

Next is the replacement of the feature of constant temperature value, for each zone of the oven, with the difference between the constant and the present temperature value. The feature of constant temperature value defines the value of temperature set by the operator of the oven and it is constant within a time interval. Lastly, there were identified 11 classes of events which are of the most interest to our scope and they were assigned to a numerical value from 1 to 11:

1. Flux Heater High Warning
2. Hi Warning
3. Lo Warning
4. Hi Deviation
5. PPM Level within limit
6. PPM Level has exceeded the amount set
7. High Water Temp Alarm Cool Down Loaded

8. Low Exhaust Alarm
9. Exhaust is insufficient
10. Heat Fan Fault
11. Blower Failure (Fan Fault).

On the ground that each sensor output and each event can be seen as a discrete variable that changes through time, the selected deep learning technique is the Recurrent Neural Network (RNN). The data are fed to the network in consecutive time series of length 32 and the network will predict the next (starting from the last point of the time series) 5 future events (25 min of events). Figure 6 is a depiction of the data format and it is worth to mention that events of both input and output will be specified in one-hot encoding format.
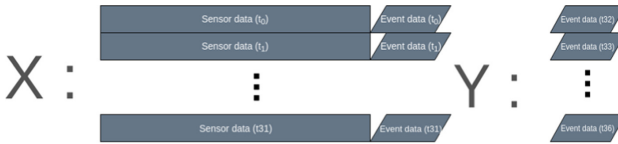

Fig. 6. Training data format.

Network Architecture. The network consists of an encoder and a decoder linked to each other. The encoder consists of the first Long Short Term Memory (LSTM) network with 256 units while the decoder is the last LSTM with 64 units. The first layer takes as input 32 time steps for the 73 features and returns only 64 vectors representing the "encoding" version of the input data. The data encoding step is repeated 5 times (the number of time step in the future we want to predict) by the Repeat Vector layer creating a new time series of $[5 \times 256]$. This newly created time series is given as input to the decoder, the second LSTM. The decoder returns the new five values along with their "sequence": this can be seen as a time series (of $[5 \times 64]$) too that passes through a Dense Layer in a "time distributed fashion". The last layer uses the softmax activation.

Results. The findings of the evaluation of the proposed method suggest that the models is able to predict future log events based on the previous ones and existing sensor data. The evaluation was made for each of the five events. The first event represents the "less in the future" class while the fifth event represents the "more in the future" class. Indicatively, in Fig. 7, the confusion matrix of the first and fifth event are depicted. The Confusion Matrix in Fig. 7 suggests that almost all the classes are predicted correctly. Note that class 4 is the most difficult to classify due to its small dimension and there are only six occurrences in the test set. Nevertheless, knowing the meaning of the class at prior the misclassification error from class 4 to class 2 is a minor error compared to the misclassification from class 4 to class 0. It is necessary to highlight the fact that the most important class is the last one, class 10, as it is the one denoting

Fig. 7. Confusion matrices for event t1 (left) and event t5 (right).

that the oven is in a fault state. Having 89% percent of correct prediction it is a considerable result taking in account the qualitative and quantitative complexity of the class.

The results of the test set are very promising, although the accuracy metric will not be taking into account due to the highly imbalanced dataset. Based on the nature of the problem, the most attention is focused on the amount of false negative results, thus, recall and f1-score are the main focus metrics and Matthews correlation coefficient is just reported. The evaluation metrics were calculated for each of the five events. The first event reached the higher evaluation score while the fifth event achieved the lowest overall score. The next table contains the proposed model results (Table 2):

Table 2. RNN evaluation metrics results.

	Recall	F1-score	MCC
Event t1	0.917	0.920	0.876
Event t2	0.876	0.882	0.818
Event t3	0.837	0.846	0.763
Event t4	0.811	0.822	0.724
Event t5	0.790	0.803	0.691

The most substantial event, the fifth one, is 25 min in the future. The evaluation results are really promising and despite the long prediction time, the performance of the classifier on the event 5 is still acceptable.

5   Conclusions

Taken together, our research has come to a conclusion towards a complete solution for sensor-based data analysis in a case-study of an industrial oven. The study was divided in two parts: the analysis of deployed acoustic sensor data

and the analysis of existing sensor data and log events. The aim of this work was to exploit operational routine for monitoring of the system performance using the already integrated sensors of the oven and then deploy extra test sensors (acoustic ones) in order to serve as indicators of a systems health condition. Both the analytic techniques achieve a failure prediction of the oven machine with high accuracy. Despite the fact that there are limitations due to the imbalanced dataset, we formed a competent technique, capable of detecting anomalies and failures on a primitive phase aiming to improve both production and maintenance efficiency. The predictive maintenance approach we presented can constitute an assisting tool for the decision support system of industries towards the prevention of potential failure and secure of safe operations of machineries.

References

1. Jardine, A.K., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Sig. Process. **20**(7), 1483–1510 (2006)
2. Sipos, R., Fradkin, D., Moerchen, F., Wang, Z.: Log-based predictive maintenance. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1867–1876. ACM (2014)
3. Hashemian, H.M., Bean, W.C.: State-of-the-art predictive maintenance techniques*. IEEE Trans. Instrum. Meas. **60**(10), 3480–3492 (2011). https://doi.org/10.1109/TIM.2009.2036347
4. Gong, C.S.A., et al.: Design and implementation of acoustic sensing system for online early fault detection in industrial fans. J. Sens. **2018** (2018)
5. Glowacz, A.: Diagnostics of DC and induction motors based on the analysis of acoustic signals. Meas. Sci. Rev. **14**, 257–262 (2014). https://doi.org/10.2478/msr-2014-0035
6. Glowacz, A., Glowacz, W., Glowacz, Z., Kozik, J.: Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. Measurement **113**, 1–9 (2018)
7. Glowacz, A.: Acoustic based fault diagnosis of three-phase induction motor. Appl. Acoust. **137**, 82–89 (2018)
8. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. In: ACM SIGMOD Record, vol. 29, pp. 93–104. ACM (2000)
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD 1996, pp. 226–231 (1996)
10. Miller, J.: Reaction time analysis with outlier exclusion: bias varies with sample size. Q. J. Exp. Psychol. **43**(4), 907–912 (1991)
11. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)