Machine Failure Prediction Using Multilabel Classification Methods

H.M.N.S. Kumari
Department of Mathematics, Faculty of Science, University of Peradeniya, Peradeniya (20400), Sri Lanka.
nadeeshaku@sci.pdn.ac.lk

U.M.M.P.K. Nawarathne
Faculty of Computing, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe (10115), Sri Lanka.
mnawarathne20@gmail.com

ABSTRACT

Early detection of machine failure is crucial in every industrial setting as it may prevent unexpected process downtimes as well as system failures. However, machine learning (ML) models are increasingly being utilized to forecast system failures in industrial maintenance, and among them, multilabel classification techniques act as efficient methods. Therefore, this study analyzed machine failure data with five types of machine failures. Initially, a feature selection approach was also carried out in this study to determine the variables which directly cause machine failure. Furthermore, multilabel k-nearest neighbours (MLkNN), multilabel adaptive resonance associative map (MLARAM), and multilabel twin support vector machine classifier (MLTSVM) in adapted algorithms, Binary Relevance, ClassifierChain, and LabelPowerSet in problem transformation approaches, and Random Label Space Partitioning with Label Powerset (RakelD) in ensemble classifiers were employed. To train these models, both the original data set as well as data frame after the feature selection was used, and hamming loss, accuracy, macro, and micro averages were calculated for each of these classifiers. According to the results, MLkNN in adapted algorithms and LabelPowerset in problem transformation approaches performed better than other classifiers used in this study. Therefore, it can be concluded that MLkNN and LabelPowerset could be used to classify multilabel with positive results.

KEYWORDS: *adapted algorithms, ensemble classifiers, feature selection, machine failure, machine learning, multilabel classification, problem transformation.*

1 INTRODUCTION

The loss of production time due to machinery breakdown is a major concern for any business that relies on manufacturing. Failure of a machine occurs when some aspect of an industrial asset does not operate as designed, leading to reduced performance or an outright shutdown. This failure of equipment can have a wide range of consequences, from insignificant to catastrophic, including increased repair costs, unscheduled downtime, lost productivity, and problems for the workers' health and safety, as well as an effect on production and the delivery of services. Machine failure can happen due to many reasons, such as operator mistakes, improper use, inadequate regular and preventative maintenance, unreliable culture, physical damage, and heating up. Therefore, it is important to predict the machine's failure in advance to reduce the unnecessary costs that may incur.

However, in recent years few studies have been done to predict machine failure using different techniques. Traditional approaches to fault diagnosis (Corne, Vervisch, Derammelaere, Knockaert, & Desmet, 2018; Glowacz et al., 2017; Irhoumah et al., 2018; Sapena-Bano et al., 2018) rely on elaborate mathematical models, including supervised diagnosis or processing system dynamic models (AntoninoDaviu & Popaleny, 2018; Bessous, Chemsa, & Sbaa, 2018; Brandt, Gutten, Koltunowicz, & Zukowski, 2018; Ullah, McDonald, Martin, Benarous, & Atkinson, 2019). With the dawn of state-of-the art technologies, industrial settings have started to employ machine learning (ML) techniques to predict the faults in machines. In order to improve the standard approach to compound-fault identification in rotating machinery, Wang, Zhang, Li, and Wu (2020) created a novel ensemble extreme learning machine (EELM) network by merging binary classifiers. They proposed an extreme learning machine (ELM) for clustering and multilabel classification and concluded that the EELM-based fault diagnosis approach provides the best overall performance through their results. Using Motor current signature analysis (MCSA)- Fourier transforms (FFT), Bessous, Sbaa, and Megherbi (2019) examined the failures in squirrel cage induction motors (SCIMs) caused by rolling element bearings (REBs). In addition, a new indication built on top of the MCSA- discrete wavelet transform (DWT) method was created, and the two methods were compared in depth. In the end, they found that MCSA-DWT provided reliable data on SCIM health.

Kankar, Sharma, and Harsha (2011) concentrated on ball-bearing fault diagnosis utilizing artificial neural networks (ANN) and support vector machines (SVM). The original vibration features were extracted, and their dimensionality was reduced using statistical approaches. From their findings, it was apparent that these ML algorithms can be employed for a fully automated bearing fault diagnosis system. Ferreira and Warzecha (2017) developed a multi-criteria framework for classifying up to ten machine conditions with a focus on experimental processes. They measured the voltages and currents in a synchronous machine. Using a sparse Linear Discriminant Analysis technique, they filtered the signals and extracted the key features they had previously identified. Scatter plots in three dimensions (with a symbol for each machine state) were used to illustrate the findings. After further examination, they determined that this technique can be applied to the diagnosis of a wide variety of machine faults.

Delgado-Arredondo et al. (2017) established a method for fault detection in induction motors in steady-state operation based on the analysis of acoustic sound and vibration signals. The signal was broken down into its component intrinsic mode functions using the complete ensemble empirical mode decomposition. In addition to identifying additional frequencies related to the defects, their proposed approach resulted in improved fault detectability outcomes compared to other published publications.

In their research, Feng, Jones, Chen, and Fang (2018) examined how various multilabel classification techniques performed in the failure classification problem. They tested eight different methods of classification on five different programs containing over eight thousand different bugs. Compared to single-label methods, the experimental results demonstrated that multilabel approaches yield higher accuracy. To determine if a specific code piece is impacted by many scents, Guggulothu and Moiz (2020) proposed and explored the usage of multilabel classification (MLC) techniques. After converting two code smell datasets from the literature into a multilabel dataset (MLD), it was discovered that the two MLC approaches took into account the association between the smells and improved performance for the 10-fold cross-validation with ten iterations.

Tan et al. (2021) analyzed the performance of different cutting-edge multilabel classification algorithms for fault diagnosis of maritime machinery using single-fault data. They used a dataset derived from a Frigate simulator that had been validated against real data to experimentally verify the efficacy of their approach. Their experiments validated the viability of the proposed approach, which can aid in making informed choices regarding the use of multilabel classification for simultaneous fault diagnosis of marine systems. In order to diagnose many defects simultaneously and assess the fault severity in noisy environments, Dineva et al. (2019) used a new method for multilabel classification. Electrical signature analysis and conventional vibration data were utilized for modelling, and the efficacy of different multilabel classification models was examined. They conducted experiments to verify the suggested method's viability under a variety of fault circumstances, including imbalance and misalignment.

The preceding summary of the literature, however, reveals that there have been relatively few studies published on the investigation of the multilabel prediction performance of contemporary classifier algorithms. In addition, there is limited interpretation when it comes to choosing the best classifier for use in the industry. Therefore, this study aims to find suitable multilabel classifiers for machine failure prediction. Section 2 of this paper discusses the materials and methods that have been used, and in section 3, the results obtained are discussed in detail. Section 4 includes the conclusion of this study.

2 MATERIALS AND METHODS

2.1 Data

This study used data related to a machine failure, and data was retrieved from an online data repository (Matzka, 2020). The original dataset is comprised of 10 000 records that describe the following machine features.

1) Product ID - Describes the product quality using the letter notation of L (50% of all products), M (medium value of 30%) and H (high values of 20%), along with a variant-specific serial number.
2) UID - A unique id to identify the products.
3) Air temperature (in Kelvin) - Air temperature was generated using a random walk process that was normalized to a standard deviation of 2 K around 300 K.
4) Process temperature (in Kelvin) - Generated by adding the air temperature plus 10 K to a random walk process with a standard deviation of 1 K.
5) Rotational speed (rotations per minute) - Calculated using a 2860 W power and a normally distributed noise as a background.
6) Torque (Newton Meters) - Torque values were considered without having negative values with a normal distribution around 40 Nm with a stand deviation of 10 Nm.
7) Tool wear (minutes) - The high-quality variations H, M, and L add 5/3, 2 minutes to the process, causing the used tool to deteriorate.

In addition to the above machine features, machine failures have been recorded considering five independent failure types as follows.

1) Tool wear failure (TWF) - The tool wear failure is recorded when a tool fails or is replaced between a time of 200-240 minutes.
2) Heat dissipation failure (HDF) - If the difference in air and process temperatures is less than 8.6 K and the tool's rotational speed is less than 1380 rpm, heat dissipation results in a process failure.
3) Power failure (PWF) - Power failures are recorded if the power is below 3500W or above 9000W. Power is the product of torque and rotational speed (in rads-1).
4) Overstrain failure (OSF) - Overstrain failures are recorded when the product of the tool wear and torque exceeds 11,000 minNm for the L-type products, 12,000 minNm for M-type products and 13,000 for H-type products, respectively.
5) Random failures (RNF) - Each process has a chance of 0.1% to fail despite the process parameters that are defined as random failures.

If at least one of the above failure modes were recorded, the machinery failure label has been recorded as ‘1’, which will indicate the malfunction of the machine. Figure 1 depicts the original form of the data frame.

UDI	Product ID	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	HDF	PWF	OSF	RNF
1	M14860	M	298.1	308.6	1551	42.8	0	0	0	0	0	0	0
2	L47181	L	298.2	308.7	1408	46.3	3	0	0	0	0	0	0
3	L47182	L	298.1	308.5	1498	49.4	5	0	0	0	0	0	0
4	L47183	L	298.2	308.6	1433	39.5	7	0	0	0	0	0	0
5	L47184	L	298.2	308.7	1408	40.0	9	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
9996	M24855	M	298.8	308.4	1604	29.5	14	0	0	0	0	0	0
9997	H39410	H	298.9	308.4	1632	31.8	17	0	0	0	0	0	0
9998	M24857	M	299.0	308.6	1645	33.4	22	0	0	0	0	0	0
9999	H39412	H	299.0	308.7	1408	48.5	25	0	0	0	0	0	0
10000	M24859	M	299.0	308.7	1500	40.2	30	0	0	0	0	0	0

Figure 1. Original data frame

2.2 Data Preprocessing and Exploration

Firstly, UDI and product Id variables were removed due to the lack of predictive power. The machinery failure variable was also removed, and TWF, HDF, PWF, OSF, and RNF were retrieved as the target columns. After that, data were checked for the availability of null values, and it was found that there were no such records. The type variable, which was originally a categorical variable, was converted to numeric values using one-hot encoding technique. However, the data were scaled using the minmaxscaler from the sklearn library (Pedregosa et al., 2011) since the column values were in different numerical ranges. After scaling, data exploration was performed to understand the data distribution. For this purpose, a correlation heat map was used, and correlation among the variables was visualized. In addition, highly correlating features that had a score greater than 0.7 were removed from the data. To be more precise about the removing variables, a feature selection was also conducted using the selectKBest method from the sklearn library (Pedregosa et al., 2011).

2.3 Multilabel Classification Techniques

Due to the inclusion of several target columns, this research problem was trained according to multilabel classification techniques. Multilabel classification techniques have the ability to provide multiple outputs compared to traditional classification methods (Herrera, Charte, Rivera, & Jesus, 2016). Firstly, the data were split so that 75% of the data were assigned for the training split while the rest of the data were allocated to the testing set. After using the train-test split approach, methodologies of problem transformation adapted algorithms, and ensemble methods were used to model the data. For this, the scikit-multilearn, which is library developed especially for handling multilabel classification tasks, was used (Szymański & Kajdanowicz, 2017).

1) Adapted algorithms – These algorithms focus on modifying cost/decision functions to adapt single-label classification algorithms to the multilabel case (Szymański & Kajdanowicz, 2017). This study implemented the multilabel k-nearest neighbours (MLkNN), multilabel adaptive resonance associative map (MLARAM), and multilabel twin support vector machine classifier (MLTSVM) for machine failure prediction.

- MLkNN – This algorithm has been developed under adapted algorithms. In MLkNN, the nearest examples to a test class are found using k-Nearest Neighbors, and assigned labels are chosen using Bayesian inference (Zhang & Zhou, 2007).
- MLARAM – This classifier approach focuses on accelerating classification by including an additional Adaptive Resonance Theory (ART) layer for grouping learned prototypes into substantial clusters. In this scenario, activating only a small portion of the prototypes can replace activating all of them, significantly reducing the classification time (Benites & Sapozhnikova, 2015).
- MLTSVM – This is a useful advancement of the twin support vector machine (TWSVM) for multilabel classification. This classifier determines multiple non-parallel hyperplanes to capture the multilabel information embedded in data (Chen, Shao, Li, & Deng, 2016).

2) Problem transformation approaches - Out of the problem transformation approaches, methods of Binary Relevance, ClassifierChain, and LabelPowerSet were utilized for training the data model in this research.

- Binary Relevance - Using the same base classifier from the constructor, the binary relevance technique divides an L-label multilabel classification problem into L separate L-label binary classification problems (Szymański & Kajdanowicz, 2017). The output of the prediction is the union of all classifiers for each label.
- ClassifierChain - This algorithm (Read, Pfahringer, Holmes, & Frank, 2009) treats each label as a link in a conditioned chain of problems involving single-class classification (Szymański & Kajdanowicz, 2017).
- LabelPowerset - In this approach to multilabel classification, a multilabel problem is transformed into a multi-class problem using a single multi-class classifier that has been trained on all unique label combinations found in the training data (Szymański & Kajdanowicz, 2017).

3) Ensemble classifiers - The application of ensemble classification schemes by ensembles of classifiers results in the generation of an array of multilabel base classifiers. In this study, only Random Label Space Partitioning with Label Powerset (RakelD) was applied. Tsoumakas, Katakis, and Vlahavas (2011) introduced RakelD as a library that has been created using an ensemble of classifiers.

However, to observe whether there is an impact of the feature selection on the classification techniques, the models were trained using both the original data frame and the data frame after the feature selection was performed.

2.4 Model Evaluation Metrics

Unlike the traditional approaches of binary classification and multi-class classification, multilabel classification has separate evaluatory metrics (Tsoumakas & Katakis, 2007). In this section, the metrics used to evaluate the results that were obtained are discussed.

1) Hamming loss – Hamming loss that is given in Eq. (1) provides a fraction of labels that are incorrectly classified, which is used to evaluate the multilabel classification methods(Ganda & Buch, 2018).

$$Hamming\ loss = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i \,\Delta\, Z_i}{M}\right| \tag{1}$$

2) Accuracy - The percentage of predicted correct labels to the total number of labels (predicted and actual) for each instance is known as accuracy, and it can be calculated as in Eq. (2) (Ganda & Buch, 2018).

$$Accuracy = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i \cap Z_i}{Y_i \cup Z_i}\right| \tag{2}$$

3) Macro average and micro average - Generally, the receiver operating characteristic curve (ROC)-area under the curve (AUC) score is generated by calculating ROC-AUC from the prediction scores. The ROC curve is a graphical method for evaluating a test's ability to distinguish between labels (Akobeng, 2007). The ROC curve can be created by calculating the test's sensitivity and specificity at every possible cut-off point and then plotting those results against 1-specificity (Akobeng, 2007). A ROC curve can also be considered the average of a test's sensitivity over all feasible specificity values or vice versa (Mandrekar, 2010). In macro ROC-AUC, for each label, it computes the metrics and determines the unweighted mean. Label imbalance is not taken into account in this. The micro ROCAUC score considers each label in the label indicator matrix when calculating metrics on a global scale.

In this analysis, both the macro ROC-AUC score and micro ROC-AUC score metrics were tested using the one versus rest method.

3 RESULTS AND DISCUSSION

After training the data models using the methods described in section 2.3, the results were recorded considering the standards of hamming loss, accuracy, macro average, and micro average. The results obtained for the original data model before applying feature selection are depicted in Table 1.

Table 1. Multilabel Evaluation Metric Scores for the Classifiers Before Applying Feature Selection

Approach type	Classifier	Hamming loss	Accuracy	Macro average	Micro average
Adaptation approach	MLkNN	0.0052	0.973	0.8826	0.9700
	MLARAM	0.2020	0.019	0.5620	0.5712
	MLTSVM	0.0078	0.964	0.5620	0.5712
Problem transformation	Binary Relevance	0.0064	0.902	0.8498	0.9630
	LabelPowerset	0.0064	0.970	0.9022	0.9640
	ClassifierChain	0.0064	0.970	0.8826	0.8930
Ensemble of classifiers	RakelD	0.0064	0.970	0.8328	0.9602

Other than the MLARAM model, the rest of the models have scored very low values for the hamming loss. From the adaptation approach, the MLkNN algorithm has the lowest hamming loss and higher scores for accuracy, macro average, and micro average. The LabelPowerset has the maximum values for evaluation metrics among the problem transformation methods. In addition, RakelD has low scores for macro and micro averages compared to MLkNN and LabelPowerset methods.

However, the multicollinearity among variables could not be overlooked when training these models. Therefore, a correlation heat map was also generated, as in Figure 2, and the features with a correlation which is greater than 0.7 were eliminated. According to the heat map, process temperature and torque variables are highly correlated with air temperature and rotational speed, respectively. To be more accurate about the dropping variables, a feature selection was also conducted using selectKBest method, and the results generated by this method also confirmed that process temperature and torque columns should be dropped from the data frame. Therefore, process temperature and torque columns were eliminated.

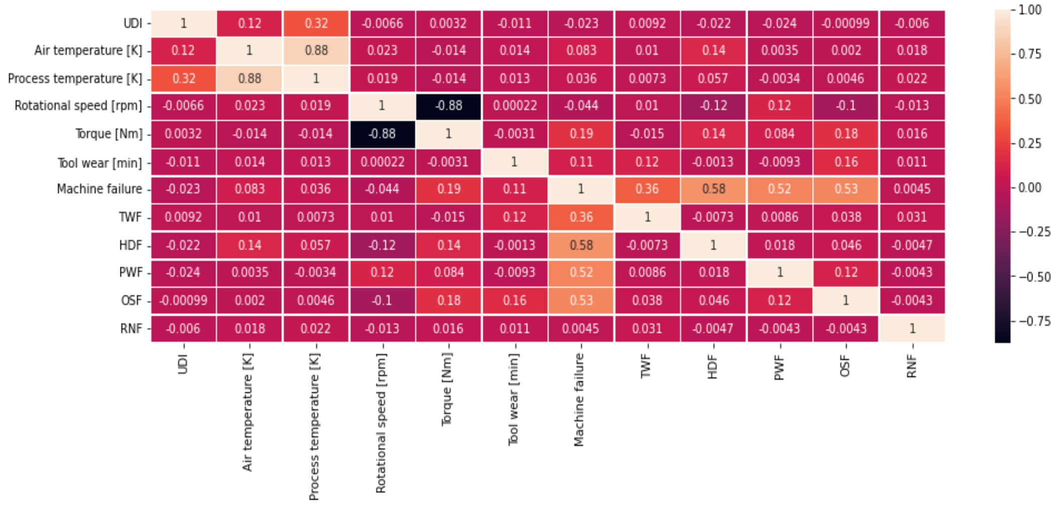

Figure 2. Correlation heatmap describing the relationship among variables

After removing process temperature and torque from the data set, the methods discussed in section 2.3 were reapplied, and the evaluation metrics were also calculated. Table 2 shows the results recorded for the respective metrics for each classification technique.

Table 2. Multilabel Evaluation Metric Scores for the Classifiers After Applying Feature Selection

Approach type	**Classifier**	**Hamming loss**	**Accuracy**	**Macro average**	**Micro average**
Adaptation approach	MLkNN	0.0050	0.970	0.8800	0.9700
	MLARAM	0.0190	0.202	0.5620	0.5712
	MLTSVM	0.0078	0.964	0.5620	0.5712
Problem transformation	Binary Relevance	0.0064	0.970	0.8400	0.9670
	LabelPowerset	0.0064	0.970	0.9000	0.9700
	ClassifierChain	0.0064	0.970	0.8800	0.9690
Ensemble of classifiers	RakelD	0.0064	0.970	0.8300	0.9692

According to Table 2, it is clear that the MLARAM has a large hamming loss value and low scores for accuracy, macro average, and micro average. MLkNN method has attained the highest metric scores among the adaption techniques, even after eliminating two variables. Out of the problem transformation methods, the LabelPowerset has gained a low hamming loss score and high values for the other metrics. Even after applying feature selection, the scores for RakelD's macro and micro averages are low in comparison to those of the MLkNN and LabelPowerset methods.

When the models were trained without applying feature selection, the scores recorded for MLkNN and LabelPowerset had optimal values. However, according to the results illustrated in Table 1 and Table 2, it can be seen that the results for both these classifiers that were trained without applying feature selection and trained after applying feature selection have similar metric values indicating negligible difference in performances.

It was also noted that, during the feature selection phase, the removed torque column could be an essential feature when being considered from the perspective of machinery parts. As power is the product between torque and rotational speed, removing torque might directly affect the predictions, especially regarding power failures. Therefore, considering the features of the machinery, it is recommended to perform feature selection, being mindful of this point.

Matzka (2020) has presented an explainable model and an explanatory interface using the original dataset used in this research work. In this study, the researcher has used explainable decision trees as well as normalized feature deviation as an explanatory interface. However, Matzka (2020) has found that in some circumstances, the decision trees offer no beneficial insights, while the normalized feature deviations offer explanations of low quality. In order to overcome these issues, our study focused on predicting machinery failures using multilabel classification techniques to provide early insights. These techniques are believed to be efficient since they offer users with possible failure type combinations, as opposed to predicting a failure without specifying which failure mode will occur.

4 CONCLUSION

In this research work, a multilabel classification approach was used to predict machinery failures. The original data set has five types of machine failures, and if at least one failure mode was recorded, the machine displayed a tendency to break down. This study performed a feature selection procedure to examine the variables which directly affect machine failure. Furthermore, seven multilabel classifiers were implemented using the original data set as well as the new data frame which was formed after applying feature selection. Hamming loss, accuracy, macro average, and micro average were calculated for each of these models in order to evaluate the performance. From the adapted algorithm approaches, MLkNN, MLARAM, and MLTSVM classifiers were used to train the data, where the MLkNN classifier performed better than the other two methods. The Binary Relevance, LabelPowerset, and ClassifierChain were used respectively from the problem transformation methods, where the LabelPowerset-based model produced substantially better results during the training phase. The RakelD classifier was selected from the ensemble of classifiers since it yielded the best results but performed poorly compared to the MLkNN and LabelPowerset classifiers. However, it is noted that feature selection did not significantly alter the scores obtained from evaluation metrics before and after they were applied. Even though the features of torque and process temperature were removed during the feature selection phase, there is a possibility for this to affect predictions considering the machinery state. Therefore, this study concludes the results with metric scores obtained before applying the feature selection. For future research, the techniques such as multilabel embeddings and label space clusters can be used to observe and compare the results. Alternatively, this study could be conducted by considering only the machinery failure column, which could be converted to a multi-class classification problem rather than a multilabel classification, and disciplines such as deep learning techniques can be utilized.

REFERENCES

Akobeng, A. K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatrica, 96(5), 644–647. https://doi.org/https://doi.org/10.1111/j.1651-2227.2006.00178.x

Antonino-Daviu, J., & Popaleny, P. (2018). Detection of Induction Motor Coupling Unbalanced and Misalignment via Advanced Transient Current Signature Analysis. 2018 XIII International Conference on Electrical Machines (ICEM), 2359–2364. https://doi.org/10.1109/ICELMACH.2018.8506949

Benites, F., & Sapozhnikova, E. (2015). HARAM: A Hierarchical ARAM Neural Network for LargeScale Text Classification. 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 847–854. https://doi.org/10.1109/ICDMW.2015.14

Bessous, N., Chemsa, A., & Sbaa, S. (2018). New Vision about the Mixed Eccentricity Fault Causes in Induction Motors and its relationship with the Rolling Element Bearing Faults : Analytical model dedicated to the REB faults. 2018 International Conference on Communications and Electrical Engineering (ICCEE), 1–11. https://doi.org/10.1109/CCEE.2018.8634566

Bessous, N., Sbaa, S., & Megherbi, A. C. (2019). Mechanical fault detection in rotating electrical machines using MCSA-FFT and MCSA-DWT techniques [Artykuły / Articles]. Bulletin of the Polish Academy of Sciences: Technical Sciences, 67(No. 3), 571–582. https://doi.org/10.24425/bpasts.2019.129655

Brandt, M., Gutten, M., Koltunowicz, T., & Zukowski, P. (2018). Analysis of winding fault in electric machines by frequency method. 2018 ELEKTRO, 1–4. https://doi.org/10.1109/ELEKTRO.2018.8398298

Chen, W.-J., Shao, Y.-H., Li, C.-N., & Deng, N.-Y. (2016). MLTSVM: A novel twin support vector machine to multilabel learning. Pattern Recognition, 52, 61–74. https://doi.org/https://doi.org/10.1016/j.patcog.2015.10.008

Corne, B., Vervisch, B., Derammelaere, S., Knockaert, J., & Desmet, J. (2018). Emulating single point bearing faults with the use of an active magnetic bearing. IET Science, Measurement & Technology, 12(1), 39–48. https://doi.org/https://doi.org/10.1049/iet-smt.2017.0128

Delgado-Arredondo, P. A., Morinigo-Sotelo, D., Osornio-Rios, R. A., Avina-Cervantes, J. G., RostroGonzalez, H., & de Jesus Romero-Troncoso, R. (2017). Methodology for fault detection in induction motors via sound and vibration signals. Mechanical Systems and Signal Processing, 83, 568–589. https://doi.org/https://doi.org/10.1016/j.ymssp.2016.06.032

Dineva, A., Mosavi, A., Gyimesi, M., Vajda, I., Nabipour, N., & Rabczuk, T. (2019). Fault Diagnosis of Rotating Electrical Machines Using Multi-Label Classification. Applied Sciences, 9(23). https://doi.org/10.3390/app9235086

Feng, Y., Jones, J., Chen, Z., & Fang, C. (2018). An Empirical Study on Software Failure Classification with Multi-label and Problem-Transformation Techniques. 2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST), 320–330. https://doi.org/10.1109/ICST.2018.00039

Ferreira, J. G., & Warzecha, A. (2017). An application of machine learning approach to fault detection of a synchronous machine. 2017 International Symposium on Electrical Machines (SME), 1–6. https://doi.org/10.1109/ISEM.2017.7993548

Ganda, D., & Buch, R. (2018). A survey on multi label classification. Recent Trends in Programming Languages, 5(1), 19-23.

Glowacz, A., Glowacz, W., Glowacz, Z., Kozik, J., Gutten, M. H., Korenciak, D., Khan, Z. F., Irfan, M., & Carletti, E. (2017). Fault Diagnosis of Three Phase Induction Motor Using Current Signal, MSAF-Ratio15 and Selected Classifiers. Archives of Metallurgy and Materials, 62(No 4). https://doi.org/10.1515/amm-2017-0355

Guggulothu, T., & Moiz, S. A. (2020). Code smell detection using multilabel classification approach. Software Quality Journal, 28(3), 1063–1086. https://doi.org/10.1007/s11219-020-09498-y

Herrera, F., Charte, F., Rivera, A. J., & del Jesus, M. J. (2016). Multilabel Classification. Springer International Publishing. https://doi.org/10.1007/978-3-319-41111-8

Irhoumah, M., Pusca, R., Lefevre, E., Mercier, D., Romary, R., & Demian, C. (2018). Information Fusion With Belief Functions for Detection of Interturn Short-Circuit Faults in Electrical Machines Using External Flux Sensors. IEEE Transactions on Industrial Electronics, 65(3), 2642–2652. https://doi.org/10.1109/TIE.2017.2745408

Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using machine learning methods. Expert Systems with Applications, 38(3), 1876–1886. https://doi.org/https://doi.org/10.1016/j.eswa.2010.07.119

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology, 5(9), 1315–1316. https://doi.org/https://doi.org/10.1097/JTO.0b013e3181ec173d

Matzka, S. (2020). Explainable Artificial Intelligence for Predictive Maintenance Applications. In IEEE (Ed.), Explainable Artificial Intelligence for Predictive Maintenance Applications (pp. 69–74). https://ieeexplore.ieee.org/document/9253083

Matzka, S. (2020, August 30). AI4I 2020 Predictive Maintenance Dataset Data Set. UCI Machine Learning Repository: Ai4i 2020 predictive maintenance dataset data set. Retrieved from https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multilabel classification. Machine Learning, 85(3), 333–359. https://doi.org/10.1007/s10994-011-5256-5

Sapena-Bano, A., Martinez-Roman, J., Puche-Panadero, R., Pineda-Sanchez, M., Perez-Cruz, J., & Riera-Guasp, M. (2018). Induction machine model with space harmonics for fault diagnosis based on the convolution theorem. International Journal of Electrical Power & Energy Systems, 100, 463–481. https://doi.org/https://doi.org/10.1016/j.ijepes.2018.03.001

Szymański, P., & Kajdanowicz, T. (2017). A scikit-based Python environment for performing multilabel classification. arXiv. https://doi.org/10.48550/ARXIV.1702.01460

Tan, Y., Zhang, J., Tian, H., Jiang, D., Guo, L., Wang, G., & Lin, Y. (2021). Multilabel classification for simultaneous fault diagnosis of marine machinery: A comparative study. Ocean Engineering, 239, 109723. https://doi.org/https://doi.org/10.1016/j.oceaneng.2021.109723

Tsoumakas, G., & Katakis, I. (2007). Multilabel classification: An overview. International Journal of Data Warehousing and Mining, 3(3), 1-13. https://doi.org/10.4018/jdwm.2007070101

Tsoumakas, G., Katakis, I., & Vlahavas, I. (2011). Random k-Labelsets for Multilabel Classification. IEEE Transactions on Knowledge and Data Engineering, 23(7), 1079–1089. https://doi.org/10.1109/TKDE.2010.164

Ullah, S., McDonald, S. P., Martin, R., Benarous, M., & Atkinson, G. J. (2019). A Permanent Magnet Assist, Segmented Rotor, Switched Reluctance Drive for Fault Tolerant Aerospace Applications. IEEE Transactions on Industry Applications, 55(1), 298–305. https://doi.org/10.1109/TIA.2018.2864718

Wang, X.-B., Zhang, X., Li, Z., & Wu, J. (2020). Ensemble Extreme Learning Machines for Compound-Fault Diagnosis of Rotating Machinery. Know.-Based Syst. 188(C). https://doi.org/10.1016/j.knosys.2019.105012

Zhang, M.-L., & Zhou, Z.-H. (2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 40(7), 2038–2048. https://doi.org/https://doi.org/10.1016/j.patcog.2006.12.019