Predictive Maintenance based on Log Analysis: A Systematic Review

Luís M. Barata[1][0000−0002−6471−4681], Sérgio Sequeira[1][0000−0001−6441−7882], and Eurico Lopes[1,2][0000−0002−1818−8203]

[1] Instituto Politécnico de Castelo Branco, Portugal
{luis.barata,ssequeira}@ipcb.pt
[2] Instituto Politécnico de Castelo Branco - Unidade de Investigação QRural, Portugal eurico@ipcb.pt

Abstract. In today's industries, the Maintenance process of machines and assets implies a significant part of the total operating cost. Many efforts have been made to reduce this cost by optimizing the process and evolving methods that allow information collection on equipment status, avoiding redundant interventions, and predicting the exact moment to perform a maintenance intervention. Using "intelligent" systems that collect data from the operation and remote management systems allows us to gather all the data and apply some methodologies capable of identifying expected behaviors based on past operations.
We present a survey of technologies, techniques, and methodologies to give the knowledge background to develop a framework to minimize the occurrence of failures and optimize the process of Predictive Maintenance (PdM) based on the analysis of Log files collected from the various industrial equipment. Generally, these logs contain many records, and many of these records do not directly contribute to evaluating the operation's machine status.
Most of the studies included in this survey use machine learning techniques and focus a significant part of their research on data preprocessing, uniformization and clarification.

Keywords: Predictive Maintenance · Log analysis · Log file · Predictive algorithms · Predictive Maintenance based on Log Analysis.

1 Introduction

We are in the era of the Fourth Industrial Revolution (Industry 4.0), also referred to as the "smart factory" [9]. Industry 4.0 advocates greater digitization through integrating information technology and industrialization in manufacturing and an intelligentization of industrial equipment [17]. Organizations face substantial challenges in the process of adapting existing processes and equipment. Industry 4.0 represents the current trend of automation technologies in the manufacturing industry, and it mainly includes enabling technologies such as cyber-physical systems (CPS), the Internet of Things (IoT), cloud computing [16], Big Data, and artificial vision [9].

All these factors also open doors for the entry of information technologies into the production and operational process, giving rise to intelligent factories, intelligent production, and intelligent operation [6]. Industries are now present with keys (data, technology, and analytics) to boost their performance in terms of operational, economics, process safety, environment, and Maintenance [11].

Predictive Maintenance (PdM) technology is designed to help predict equipment failures so that corrective maintenance can be scheduled in advance, prevent unexpected equipment downtime, improve customer service quality, and reduce additional costs caused by excessive Maintenance of preventive maintenance strategies [14].

The Industrial Internet of Things (IIoT), a subset of the IoT (Internet of Things), is one of the essential concepts contributing to Industry 4.0. It is characterized by the combining of industrial systems with advanced computing, sensors, and ubiquitous communication systems [13]. All this number of sensors and intelligent equipment register their activity in the form of log files. This information is used to monitor the equipment's correct function and knows more about the system and how the various components perform in an industrial environment. The Log data generated is enormous, very distinct information, and scrambled together, needing to be sorted in some manner to be processed and return valid information.

If we know how a system has performed in the past, we can predict how it will perform. So this reveals the importance of applying knowledge from past failures and using this information to optimize the function of a system and the right moment when Maintenance must be performed [11].

Since most manufacturing factories already collect event data for improving productivity and control purposes, this data can be used for forecasting interventions based on historical events with minimal overheads [3].

A critical aspect is the best way to analyze the data, so this study aims to review the different algorithms that can be used in the process of Predictive Maintenance. The scope of this review is related to the importance of the most accurate algorithm to get more reliable predictions. After filtering the different databases, we examined several related papers found until 30th December 2022.

The remaining sections of the essay are structured as follows. In Section 2, the applicable methodology is defined, along with the study objectives, inclusion standards, and search approach. The findings of this systematic study are presented in Section 3 and discussed in Section 4. The paper is concluded in Section 5.

2 Methods

To perform this investigation, a systematic review approach was used based on the advantages of synthesizing all the information available methodically by using repeatable and well-defined actions to answer a clearly formulated research question(s).

2.1 Research Questions

This systematic review is based on the following questions:

(RQ1) How can Log Analysis be used in Predictive Maintenance?

(RQ2) Which algorithms can be used in Log Analysis?

(RQ3) Based on the algorithms identified, what are the benefits of implementing each one?

2.2 Inclusion Criteria

The study of Predictive Maintenance based on Log Analysis was performed with the following inclusion criteria: (1) Studies that use data log analysis to predict equipment maintenance; (2) Studies that present the methods or algorithms used in Log Analysis; (3) Studies published in conferences or journals; (4) Studies that access to full-text to be included; (5) Studies that show results; (6) Studies that are original research; (7) Studies that were published between 2012 and 2022; (8) Studies written in English.

2.3 Search Strategy

This systematic review consists of the studies that follow the inclusion criteria in the following electronic databases: IEEE Xplore and ACM Digital Library. The following research terms were used to research this systematic review: "Predictive Maintenance" AND "Log". Every study was independently evaluated by the authors, determining its suitability. The studies were analyzed to identify the different methods for using log data to generate actions for Predictive Maintenance tasks.

Figure 1 shows that a large part of the studies is focused on the last five years, indicating that this is a topic that has raised the interest of researchers probably as a result of the advent of Industry 4.0 [17], which has aroused the scientific community's awareness for reality of industrial Maintenance.

2.4 Extraction of Study Characteristics

Several parameters were extracted from the different studies. The extracted data from the various studies were presented in Table 1: year of publication, location, infrastructure or theme, and algorithms or methods used.

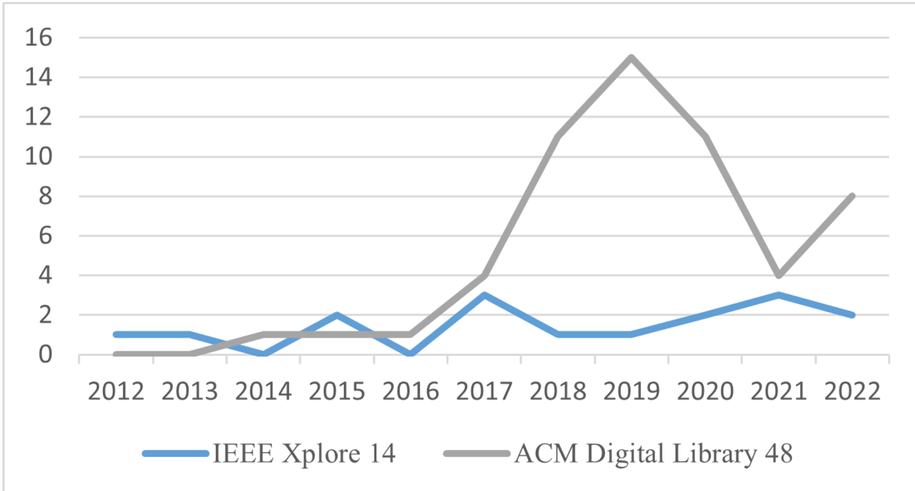

Fig. 1. Studies published from 2012 to 2022. (self)

Table 1: Summary of publications and their corresponding locations, infrastructure/themes, and algorithms or methods used.

Ref.	Location	Infrastructure/Theme	Algorithms or Methods
[1]	Oslo	Electric propulsion systems of large ships	Intelligent Predictive Maintenance (IPdM) based on Balanced Random Forest and Multiple Instance Learning
[2]	India	Diagnostic Service for IOT enabled Smart Refrigerators	Predict the error at least 3 days in advance using XGBoost (Extreme Gradient Boosting) classifier
[4]	Italy	Tier-1 data center of the Italian Institute for Nuclear Physics (INFN) - CERN	eGFC: Evolving Gaussian Fuzzy Classifier
[5]	Bologne, Italy	Tier-1 data center of the Italian Institute for Nuclear Physics (INFN) - CERN	Evolving fuzzy and neuro-fuzzy granular classifiers, FBeM and eGNN
[10]	Sweden	Predicting future Maintenance of Volvo Trucks based on diagnostic trouble codes (DTCs)	Machine Learning - Supervised learning classification-based predictive maintenance framework

Continued on next page

Table 1 – continued from previous page

Ref.	Location	Infrastructure/Theme	Algorithms or Methods
[8]	Sidney, Australia	Failure Prediction of Railway Points on Sydney Trains rail network	Multiple Kernel Learning framework
[15]	Vientiane	Nam Ngum-1 hydropower plant	Support Vector Machines and Decision Trees
[7]	France	Time-to-failure prediction in Aviation by using Post Flight Report data - AIRBUS	Combination of statistical and machine learning techniques
[12]	USA	Predictive Maintenance in Siemens AG Healthcare and Siemens Medical Solutions equipment's	Multi-Instance Learning (MIL)
[3]	Singapore, Thailand	High-performance Manufacturing Systems - Semiconductor Company	Statistical multiple Regression Forecasting, Backpropagation Neural Network (BPNN), and Evolvable NN (ENN) based on Genetic Algorithm (GA)

3 Results

In the review process, shown in 2, we picked 72 publications from the databases that were available, and we found that six of the publications were identical to one another. After examining the titles, abstracts, and keywords of all research articles, we determined that 46 of the studies could not be included in the analysis because they did not have a direct bearing on the use of log analysis to predict the need for Maintenance. By reading the introduction and final of the publications, five were discarded for not refereeing algorithms or methods used. The entire texts of the remaining 15 papers were reviewed considering the inclusion criteria, and as a result, five articles were disqualified from further consideration. In the end, the 10 articles that remained after the first review were looked at, and their results were included into both the qualitative and quantitative synthesis.

The studies were reviewed, and a selection was made, after which the pertinent material and metadata were extracted. The research that was carried out for this study discovered research publications that were published between the years 2012 and 2022. 80% of the studies were published in the last five years (2018 to 2022), with 20% published in 2018, 2019 and 2020, and 10% published in the last two years. In terms of the methodologies and algorithms that are used, 90% make use of Machine Learning procedures, while 10% make use of a combination of Machine Learning and Statistical Methods. All the studies are performed in well-known applications and organizations, so this is an issue that drives attention to most of the organizations of reference.

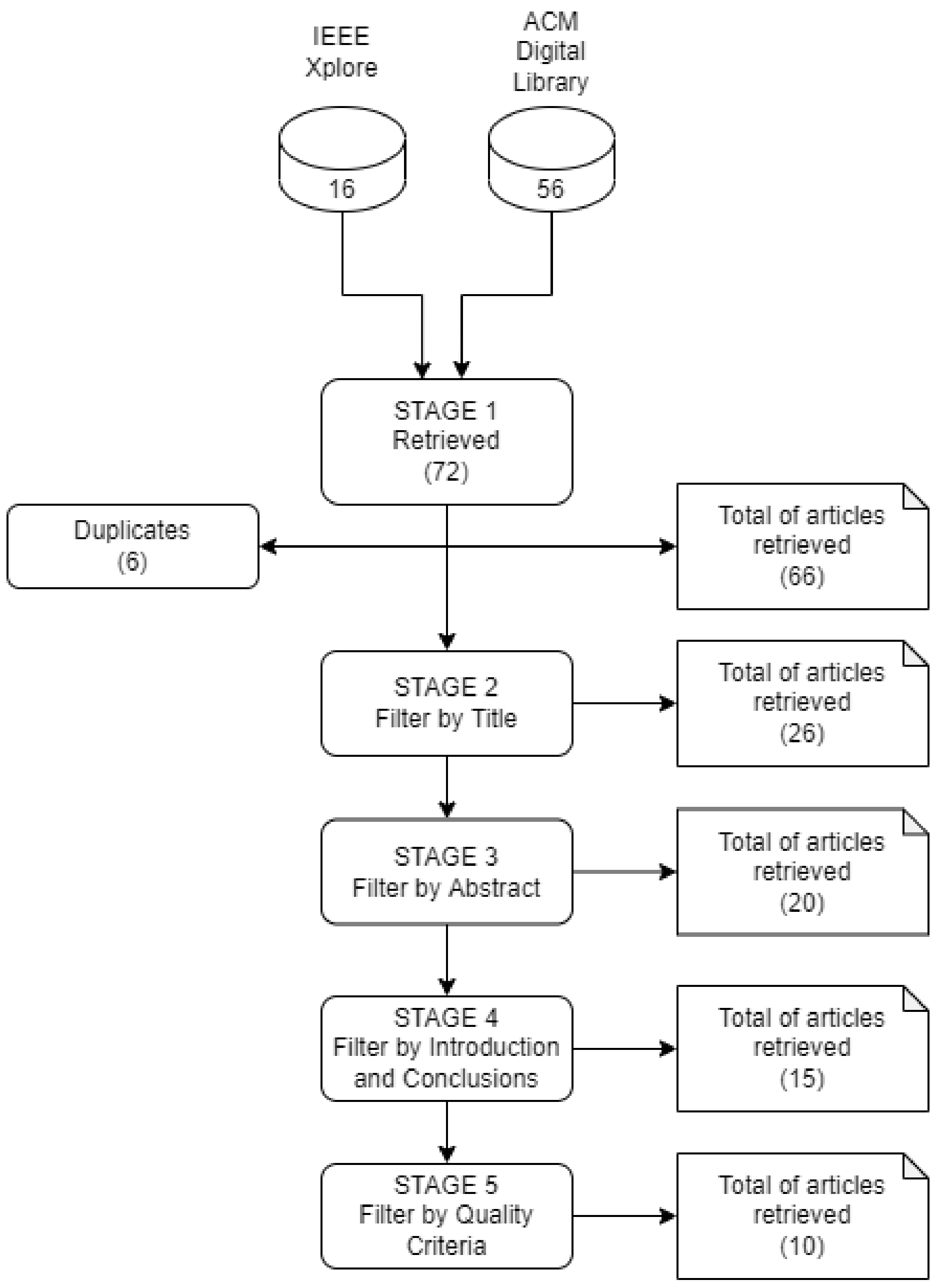

Fig. 2. Flow diagram of the selection of the papers. (self)

4 Discussion

In [3], the authors identify the significant challenges faced when processing data from log files as follows:

- Lack of an appropriate data processing approach;
- Unknown correlations between various parameter values; and
- Noisy and erroneous data acquired during production and maintenance.

This facts are critical for accurate reporting and warning of equipment failure before it occurs. To obtain higher fault prediction accuracy, the task of interpreting industrial raw data must be clearly defined. Under this framework, three techniques can be used, multiple statistical Regression Forecasting, Backpropagation Neural Network (BPNN), and Evolvable NN (ENN) based on a Genetic Algorithm (GA) to predict the health of equipment and tested based on historical data.

Korvesis et al. proposed a method to solve the problem of event prediction, and the purpose is to perform predictive Maintenance in aviation. Given a set of recorded events corresponding to equipment failures, their method will predict the next occurrence of one or more related events (target events or severe shortcomings) [7].

Considering the occurrence of other events in the past, they formulated a regression problem to approximate the risk of the target event. They adopted a multi-instance learning scheme (multi-instance regression) and extensive data preprocessing functions to obtain the best results. The method was applied to event logs generated by monitoring systems onboard the aircraft, registering events corresponding to component failures. The primary purpose was to predict those failures that could be critical to the function of the plane. Anticipating these problems beforehand and scheduling the necessary Maintenance boosts the aircraft's availability because when one of these issues happens, the aircraft must stay on the ground for control or repair. The study also shows that predictive Maintenance requires minimal false positives. Their incidence and timing relative to the goal event must be studied.

By the other hand, some authors choose to predict the maintenance time and recommend using the recorded data and applying Machine Learning technology to the preventive Maintenance of the cooling system in power plants. The system's historical data and maintenance details are prepared for the Machine Learning algorithm, classify the data using a Classification Learner Application, so this study greatly emphasizes the task of data preparation [15]. The Classification Learner Application provides support for a total of 22 different classifier types. These classifier types are arranged into six different major classification algorithms, which are as follows: Decision Trees, Discriminant Analysis, Support Vector Machines (SVM), Logistic Regression, k-Nearest Neighbors (KNN), and Ensemble Classification.

This program uses supervised machine learning to classify data. Data may be explored, features selected, validation systems specified, models trained, and outcomes assessed. The program automatically searches for the optimum classification model type, including Decision Trees, Discriminant Analysis, Support Vector Machines, k-Nearest Neighbors, and Ensemble Classification. This application includes confusion matrices (CM) and ROC curves. In this author's experimentation, SVM and Decision Trees are better at predicting results than the other algorithms.

In another publication, the authors recognize Log-based predictive Maintenance of computing centers as a primary concern regarding the worldwide computing grid that supports the CERN (European Organization for Nuclear Research) physics experiments. Log information preparation is a tedious and lengthy computational task, and the objective is to get essential data from a constantly inconsistent framework to develop a classification model. Evolving granular classifiers are fit to gain from time-varying log streams and, consequently, perform online classification of the severity of anomalies. They used Fuzzy-set-Based evolving Modeling (FBeM) and evolving Granular Neural Networks (eGNN) to model and monitor logging activity rates. These models were created from time-window data streams. FBeM and eGNN use incremental learning algorithms to adjust their parameters and structure. As time periods with abnormalities are identified, Predictive Maintenance can focus on those logs. Classification accuracy, model compactness, and processing time are compared [5].

Sipos et al. represent the Predictive Maintenance Workflow and point out some challenges in getting predictive data from logs. Since logs are mainly utilized for troubleshooting purposes, they (i) rarely contain unequivocal data for failure prediction; (ii) contain heterogeneous data including symbolic sequence, numeric time series, categorical variables, and unstructured text; and (iii) contain huge amounts of information, presenting computational difficulties. To utilize log information, there's the need to decipher the logs initially, sift through a lot of noise (for example, information irrelevant to the objective), and extract predictive features. Next, there's the need to collect known failure cases for learning/evaluating models, transform the error into a suitable learning situation, and decide on a performance measure that reflects real-world needs. At that point, it's time to apply machine learning techniques to take care of the learning problem effectively and efficiently. Also, there's the need to consider the errors associated with the business specifications [12].

The authors of [10] state that, Machine learning is perhaps the most encouraging methodology for tending to the difficulties that emerge in Predictive Maintenance. It has been demonstrated that machine learning, given an adequate sum and quality data, can prompt exact failure prediction and find examples of patterns of interest from sensor measurements.

The fundamental purpose behind those triumphs is the ability of algorithms, for example, artificial neural networks, random forests, or support vector machines, to find patterns that are too complex for human specialists to catch. Notwithstanding, those techniques are still heavily dependent on careful data cleaning and feature extraction, making it particularly critical to plan interactive solutions where specialists share their insight with a framework continuously and can provide direct feedback.

To predict the future Maintenance of Volvo Trucks based on diagnostic trouble codes, they developed a Machine Learning framework using Supervised learning classification-based. They collect data from 30.000 trucks that record DTCs in the vehicle's electrical control units while driving. Later, they are loaded into a central database when visiting workshops, enabling information to make the model grow.

Other authors used a Multiple Kernel Learning (MKL) framework to predict Railway Point's failure on the Sydney Trains rail network. Unlike deep neural networks, MKL enjoys better interpretability while requiring less training data, which is more in line with our fundamental requirements. MKL searches for an optimal combination of kernel functions to maximize a generalized performance measure. It has been widely used in various regression and classification tasks. They feed the model with railway points equipment details, maintenance logs, movement logs, and failure history from the Sydney Trains database between 2014 and 2017. The weather data from the Australia Bureau of Meteorology of the same time span was also given to the model to integrate all related information [8].

Decker et al. define Log records as unstructured data, so log information data should be prior prepared by learning and modeling algorithms. Using general-purpose solutions based on the content of log files has been a challenge over the years. The information might be exceptionally lengthy in a log-based framework, so extracting useful information from raw data is hard. The amount of data is enormous, while a high rate will generally be repetitive and redundant. Any computing center service run by a user generates log data using multiple files. After being processed, a reasonable amount of data for analysis is obtained. An eGFC: Evolving Gaussian Fuzzy Classifier was used in Tier-1 data center of the Italian Institute for Nuclear Physics (INFN) – CERN, which has approximately 40,000 CPU cores, 40 PB of disk storage, and 90 PB of tape storage, so the amount of log data is hudge [4].

Another application of log analysis for PdM uses information from smart refrigerators. Being IoT enabled, home appliance manufacturers are increasingly equipping their products with a series of sensors capable of recording the operation of the equipment, as well as recording and analysis capabilities. In the study by Bansal et al. (2021), this is shown by the fact that home freezers communicate event logs to a centralized system. Each log comprises data from numerous sensors, their timestamps, and some static information like the area, model code, and device identification string (Device ID). These logs are sparse, with each log including information for just a few characteristics (only sensor values that have changed are recorded), some of which may include inaccurate information due to malfunctioning sensors or network problems. This method seeks to forecast the fault at least three days beforehand so that the user may either be informed of the problem and given advice on how to fix it or a planned engineer visit can be made to address the problem right away. Extreme Gradient Boosting, or

XGBoost, is the machine learning method employed for this project. Gradient boosting trees are implemented in XGBoost, an open-source program [2].

Bakdi et al. analyze the topic of diverging data and offers a solution for intermittent, unbalanced, and unlabeled data problems when traditional approaches are no longer useful. Using the Event Logs from ship Electric Propulsion Systems, Intelligent Predictive Maintenance (IPdM) based on Balanced Random Forest and Multiple Instance Learning was created. First, irregular occurrences include alerts, warnings, and operational information from numerous units and control systems. Second, a few failures and various failure types cause imbalanced datasets and biased forecasts. Third, training datasets are inadequately labeled; only the failure date is known, not earlier causes or early symptoms. Temporal Random Indexing transforms log messages into a lower-dimensional numerical space for time-series studies. Random-forest models are balanced for unbiased classification and regression [1].

The information in this section is compiled in table 2, where algorithms and methods are presented, the performance published by the authors is collected, and the major conclusions or benefits of each technique.

Table 2: Algorithms, Performance, and Conclusions

Algorithms or Methods	Pub.	Performance	Benefits/Conclusions
MIL: Multiple Instance Regression	[7]	64% Precision; 23% Recall; 34% F1-Score	Data preprocessing is an extensive task, and data is divided into episodes, corresponding to the period between occurrences or the time of usage of a component. MIL handles with success with imbalances of the features data.
DA: Discriminant Analysis	[15]	91.3% - 99.2% Accuracy	This method accuracy was the lower, but it enables a very fast training process, and the prediction speed is high.
DT: Decision Trees	[15]	100% Accuracy	DTs have a minor training time compared with SVM, with similar accuracy.
eGFC: Evolving Gaussian Fuzzy Classifier	[4]	94.2% Accuracy	A semi-supervised evolving classifier that can handle with partial labelled data. Very good when dealing with large amount of data and allows model scalability.

Continued on next page

Table 2 – continued from previous page

Algorithms or Methods	Pub.	Performance	Benefits/Conclusions
eGNN: evolving Granular Neural Network	[5]	95.96% Accuracy	No need to classified data. Doesn´t need large number of features or data, to have an accuracy of more than 85%, allowing savings in computation effort.
Ensemble Classifiers	[15]	82.7% - 100% Accuracy	Slow method in training stage, although with the best prediction speed.
FBeM: Fuzzy set-Based evolving Modeling	[5]	83.77% Accuracy	No need to classified data. Achieves is most accuracy with a larger time window (more data).
KNN: k-Nearest Neighbors	[15]	96.6% - 100% Accuracy	Slow method in training stage, although with better prediction speed than SVM and DT.
Machine Learning - Supervised learning classification-based predictive maintenance framework	[10]	N/A	Applying the method to aggregated data (data from more than one component) provides better performance than in individual data from a simple component.
MIL: Multi-Instance Learning	[12]	N/A	When a very small number of features is used the performance of the model is decreased due to noise and sparsity.
MIL-B-RF: Balanced Random Forest and Multiple Instance Learning	[1]	FPR $<$ 6.7%; TPR $<$ 94.4%	Allows focus on time-correlation between events, very important in cyclical events or occurrences. Presents a high true positive rate, so is capable on predicting most errors or negative events.
SAMLK: Sample Adaptative Multiple Kernel Learning	[8]	N/A	Multiple Kernel Learning (MKL) needs less training data than deep neural networks. Data is analyzed based on weekdays, and the information is separated by day to incorporate the different behavior of the data over the week.

Continued on next page

Table 2 – continued from previous page

Algorithms or Methods	Pub.	Performance	Benefits/Conclusions
Statistical multiple Regression Forecasting, Backpropagation Neural Network (BPNN), and Evolvable NN (ENN) based on Genetic Algorithm (GA)	[3]	84.2% Accuracy (BPNN) 81.3% Accuracy (ENN)	The complex nature of the industrial background generates a high degree of nonlinearity. The application of a two-phase method improves the global accuracy of the prediction system.
SVM: Support Vector Machines	[15]	99.7 - 100% Accuracy	Higher training time, but accuracy close to 100%
XGBoost: Extreme Gradient Boosting classifier	[2]	96.2% Accuracy; 91.4% Precision	Suitable for use in low-class imbalance, where the number of positive and negative occurrences are balanced. It can be used in high-class imbalance by using a higher penalty for positive misclassification and lower penalty for negative misclassification.

5 Conclusions

In this article, we have performed a systematic review on the use of data stored in log files to predict maintenance necessities. Only ten articles were considered relevant per the inclusion criteria, which means that this area may be an attractive field for future research. Various algorithms also show that many solutions can be used, depending on the type and quality of information. The ideal solution can be obtained with a mix of technologies and methods.

The research also shows that in the last years, many methods of intelligence computation have been applied mainly with Machine Learning techniques to solve real-world problems. But these methods didn't replace traditional statistics models; on the other hand, they complement them.

The main findings and conclusions are the following:

(RQ1) How can Log Analysis be used in Predictive Maintenance? By processing data collected in the past, one can try to anticipate future events. If we know what happened within a system, it's expected that the same behavior will occur in the future.

(RQ2) Which algorithms can be used in Log Analysis? Besides the traditional statistic algorithms, machine learning algorithms are the tendency for PdM based on Log analysis as represented in table 2. Most approaches for failure prediction attempt to solve a classification issue; given the events that occur in a specific time frame, the classifier assesses whether or not a failure will occur within a specified period.

(RQ3) Based on the algorithms identified, what are the benefits of implementing each one? Each application exhibits unique characteristics that significantly impact the design of the corresponding algorithm. Table 2 shows a resume of the main advantages and applications of the methods.

The main challenge for a successful application of an algorithm is the quality of information grabbed in log files. Noisy and inaccurate data collected can increase the difficulty of the process, and the need to identify which data is relevant is a problem that maintenance engineers must deal. The relation between data and parameters grabbed by different sensors and/or equipment is a particular task applied to each reality.

Data clarification is the most crucial task to get reliable information on which algorithms can perform the most accurate predictions.

Author contributions

Luís M. Barata: research, methodology, formal analysis, validation, preparation of original draft, revision-writing, editing and supervision.

Sérgio Sequeira: review-writing and editing.

Eurico Lopes: proofreading, editing, and supervision.

References

1. Bakdi, A., Kristensen, N.B., Stakkeland, M.: Multiple Instance Learning with Random Forest for Event-Logs Analysis and Predictive Maintenance in Ship Electric Propulsion System. IEEE Transactions on Industrial Informatics pp. 1–1 (2022). https://doi.org/10.1109/TII.2022.3144177, conference Name: IEEE Transactions on Industrial Informatics
2. Bansal, T., Agrawal, S.S., Kumar, D., Shambu, M.T., Inbarajan, P.: AI based Diagnostic Service for IOT enabled Smart Refrigerators. In: 2021 8th International Conference on Future Internet of Things and Cloud (FiCloud). pp. 163–168 (Aug 2021). https://doi.org/10.1109/FiCloud49777.2021.00031
3. Bin Hu, Pang, C.K., Luo, M., Li, X., Hian Leng Chan: A two-stage equipment predictive maintenance framework for high-performance manufacturing systems. In: 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA). pp. 1343–1348 (Jul 2012). https://doi.org/10.1109/ICIEA.2012.6360931, iSSN: 2158-2297
4. Decker, L., Leite, D., Giommi, L., Bonacorsi, D.: Real-Time Anomaly Detection in Data Centers for Log-based Predictive Maintenance using an Evolving Fuzzy-Rule-Based Approach. In: 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). pp. 1–8 (Jul 2020). https://doi.org/10.1109/FUZZ48607.2020.9177762, iSSN: 1558-4739
5. Decker, L., Leite, D., Viola, F., Bonacorsi, D.: Comparison of Evolving Granular Classifiers applied to Anomaly Detection for Predictive Maintenance in Computing Centers. In: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). pp. 1–8 (May 2020). https://doi.org/10.1109/EAIS48028.2020.9122779, iSSN: 2473-4691

6. Erol, S.: Where is the Green in Industry 4.0? or How Information Systems can play a role in creating Intelligent and Sustainable Production Systems of the Future. vol. 2 (Sep 2016)
7. Korvesis, P., Besseau, S., Vazirgiannis, M.: Predictive Maintenance in Aviation: Failure Prediction from Post-Flight Reports. In: 2018 IEEE 34th International Conference on Data Engineering (ICDE). pp. 1414–1422 (Apr 2018). https://doi.org/10.1109/ICDE.2018.00160, iSSN: 2375-026X
8. Li, Z., Zhang, J., Wu, Q., Gong, Y., Yi, J., Kirsch, C.: Sample Adaptive Multiple Kernel Learning for Failure Prediction of Railway Points. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2848–2856. KDD '19, Association for Computing Machinery, New York, NY, USA (Jul 2019). https://doi.org/10.1145/3292500.3330731, https://doi.org/10.1145/3292500.3330731
9. Oliveira, M., Afonso, D.: Industry Focused in Data Collection: How Industry 4.0 is Handled by Big Data. In: Proceedings of the 2019 2nd International Conference on Data Science and Information Technology. pp. 12–18. DSIT 2019, Association for Computing Machinery, New York, NY, USA (Jul 2019). https://doi.org/10.1145/3352411.3352414, https://doi.org/10.1145/3352411.3352414
10. Pirasteh, P., Nowaczyk, S., Pashami, S., Löwenadler, M., Thunberg, K., Ydreskog, H., Berck, P.: Interactive feature extraction for diagnostic trouble codes in predictive maintenance: A case study from automotive domain. In: Proceedings of the Workshop on Interactive Data Mining. pp. 1–10. WIDM'19, Association for Computing Machinery, New York, NY, USA (Feb 2019). https://doi.org/10.1145/3304079.3310288, https://doi.org/10.1145/3304079.3310288
11. Reis, M.S., Gins, G.: Industrial Process Monitoring in the Big Data/Industry 4.0 Era: from Detection, to Diagnosis, to Prognosis. Processes **5**(3), 35 (Sep 2017). https://doi.org/10.3390/pr5030035, https://www.mdpi.com/2227-9717/5/3/35, number: 3 Publisher: Multidisciplinary Digital Publishing Institute
12. 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. KDD '14, Association for Computing Machinery, New York, NY, USA (Aug 2014). https://doi.org/10.1145/2623330.2623340, https://doi.org/10.1145/2623330.2623340
13. Thames, L., Schaefer, D.: Software-defined Cloud Manufacturing for Industry 4.0. Procedia CIRP **52**, 12–17 (Jan 2016). https://doi.org/10.1016/j.procir.2016.07.041, http://www.sciencedirect.com/science/article/pii/S2212827116307910
14. Wang, J., Li, C., Han, S., Sarkar, S., Zhou, X.: Predictive maintenance based on event-log analysis: A case study. IBM Journal of Research and Development **61**(1), 11:121–11:132 (Jan 2017). https://doi.org/10.1147/JRD.2017.2648298, conference Name: IBM Journal of Research and Development
15. Xayyasith, S., Promwungkwa, A., Ngamsanroaj, K.: Application of Machine Learning for Predictive Maintenance Cooling System in Nam Ngum-1 Hydropower Plant. In: 2018 16th International Conference on ICT and Knowledge Engineering (ICT KE). pp. 1–5 (Nov 2018). https://doi.org/10.1109/ICTKE.2018.8612435, iSSN: 2157-099X

16. Xu, L.D., Xu, E.L., Li, L.: Industry 4.0: state of the art and future trends. International Journal of Production Research **56**(8), 2941–2962 (Apr 2018). https://doi.org/10.1080/00207543.2018.1444806, https://doi.org/10.1080/00207543.2018.1444806, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00207543.2018.1444806
17. Zhou, J.: Digitalization and intelligentization of manufacturing industry. Advances in Manufacturing **1**(1), 1–7 (Mar 2013). https://doi.org/10.1007/s40436-013-0006-5, https://doi.org/10.1007/s40436-013-0006-5