PREDICTION OF MILLING MACHINE FAILURES USING MACHINE LEARNING MODEL

[1*]Olorunfemi B. J., [1]Obisesan O. M., [2]Oginni O. T. and [1]Olumoroti I. A.

[1]*Department of Mechanical Engineering, Federal University Oye-Ekiti, Nigeria.*
[2]*Department of Mechanical Engineering, Bamidele Olumilua University of Education, Science and Technology, Ikere-Ekiti, Ekiti State, Nigeria.*

Corresponding author's email: oginni.olarewaju@bouesti.edu.ng

ABSTRACT

This study developed a random forest machine learning model for predicting milling machine failure using five input parameters, which include air temperature, process temperature, rotational speed, torque, and tool wear. The model utilized a random forest classifier to split the data into 97% for training and 3% for testing, generating final results based on majority votes. The random forest model effectively predicts milling machine failure with high precision and accuracy, as demonstrated by its performance metrics, including accuracy, precision, recall, and F1 values of 0.9853, 0.7129, 0.8276, and 0.7660, respectively. The confusion matrix analysis shows the model correctly predicted 2884 no machine failures as true positives (TP) out of 2899 no machine failure targets and 15 no machine failures as false positives (FP). In addition, the model predicted 72 machine failure targets as true negatives (TN) out of 101 machine failure targets, leaving 29 as false negatives (FN). The model aids in predicting machine failure likelihood and implementing preventative measures such as pre-emptive investigation, maintenance schedule adjustments, and repairs. This improves the efficiency and productivity of milling machine operations, reducing unplanned downtime.

Keywords: *Accuracy, Downtime, Performance, Positive, Operation*

INTRODUCTION

Maintenance is an integral part of the production strategy for the overall success of an organization. It is expected that equipment of this century should be more computerized and reliable, in addition to being vastly more complex. Further computerization of equipment would significantly increase the importance of software maintenance, approaching, if not equal to, hardware maintenance (Gala *et al.*, 2016; Achour *et al.*, 2017). This century sees more emphasis on maintenance with respect to such areas as the human factor, quality, safety, and cost-effectiveness. New thinking and new strategies are required to realize potential benefits and turn them into profitability. All in all, profitable operations will be the ones that have employed modern thinking to evolve an equipment management strategy that takes effective advantage of new information, technology, and methods (Herath *et al.*, 2021).

Predictive maintenance (PM) is a method to monitor the status of machinery to prevent expensive failures from occurring and to perform maintenance when it is required. From visual inspection, which is the oldest method, PM has evolved to automated methods that use advanced signal processing techniques (Benmouiza and Cheknane, 2013). Traditionally, maintenance creates a trade-off situation in which one must choose between maximizing the useful life of a part at the risk of machine downtime (run-to-failure) and maximizing up-time through early replacement of potentially good parts (time-based PM), which has been demonstrated to be ineffective for most equipment components considered flawed and unreliable in recent years (Rahimikhoob, 2010; Chen and Li, 2014). PM breaks these tradeoffs by empowering companies to minimize maintenance and forecasting it ahead of time. Adoption of PM allows for the maximization of the useful life of assets by reducing the frequency of maintenance activities, avoiding unplanned breakdowns, and eliminating unnecessary preventive maintenance. This results in substantial time and cost savings and higher system reliability. To implement a PM approach, a condition monitoring (CM) system is necessary. Using the words of Chen *et al*. (2013), CM is "the process of monitoring one or more parameters of a machine to predict its potential faults early." A TCM could prevent tool wear, allow optimum utilization of the tool life, and improve the efficiency of the machine. TCM has a foremost importance in metal cutting due to its direct impact on the quality of the machined surface, its dimensional accuracy, and, consequently, the economics of machining. Even if the tool is not broken yet, its degradation reduces the work surface quality and leads to a significant loss of dimensional accuracy (Grincova and Marasova, 2014). On the other hand, excessive preventive replacement of tools involves higher costs and production time; it will require additional tools to be purchased, which are generally expensive, as well as considerable time to change the tool. The Internet of Things (IoT) enabled the presence of abundant sensors that, in real-time, collect big data composed of time domain features. The current technologies are so developed that the scientific community is no longer studying how to detect manufacturing data but which method is the most economical one (Grinˇcov´a and MArAsov´a 2014; Torabi *et al.*, 2018). PM is performed with machine learning (ML) methods that are much more accurate and can take into account all the factors provided by the sensors. With the advent of the IoT and machine learning methods, manufacturing systems can monitor physical processes and make smart decisions through real-time communication and cooperation with humans, machines, sensors, and so forth (Tienbui *et al*., 2019). Artificial intelligence enables manufacturers to reduce equipment downtime, spot production defects, improve the supply chain, and shorten design times by using machine learning technologies that learn from experience. One of the last applications of these technologies is the development of predictive maintenance systems (Park *et al.*, 2015; Paller and Elo, 2022).

The untimely and sudden downtime of machine-on-operation modules, specifically in mass production, and urgent needs called for more research. Accurate and precise breakdown times have not been successfully predicted. The study on the prediction of milling machine failure using machine learning models focused on developing a predictive model that can identify potential failures or malfunctions in generic algorithm milling machines. This development would improve the efficiency and productivity of milling machine operations by enabling proactive maintenance and reducing unplanned downtime.

MATERIALS AND METHOD

Data Collection

The machine learning model was developed using a milling machine set collected at the Project Development Institute (PRODA), Enugu State. A random forest model was built using five input parameters (air temperature, process temperature, rotational speed, torque, and tool wear). The model was developed using a random forest classifier pipeline with 100 estimators of Bernoulli Naive Bayes and a learning rate of 0.1 as a minimum condition. The data set was split into two sets for training and testing the random forest model. Seventy percent (97%) of the 3000 data set was used for sample training, while the remaining 3% was used for model testing.

Description of the Random Forest Model

Random forest is a machine learning algorithm that builds multiple decision trees and merges them to create a forest for more stable and accurate prediction. The multiple decision trees are called estimators, and each tree in the ensemble is comprised of a data sample drawn from a training data set with replacement, called the bootstrap, as shown in Figure 1 (Diez-Olivan *et al.*, 2017). It is deployed for both classification and regression problems. A random forest classifier collects the result of each decision tree and expects the final output based on the majority votes of predictions (Jansen *et al*., 2018; Grincova and Marasova, 2014).

Equation 1 is used to make and calculate a prediction at a new point x classification according to Archour *et al.*, 2017.

$$C^B(x)=\text{majority vote(mean)}\{C_b(x)^B\} \quad (1)$$

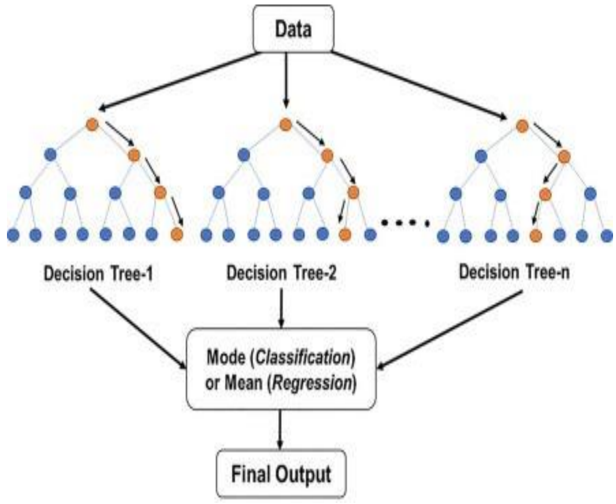

Figure 1: Decision forest

where C_b (x) is the class prediction of both random forests. The flow chart of the random forest algorithm is presented in Figure 2

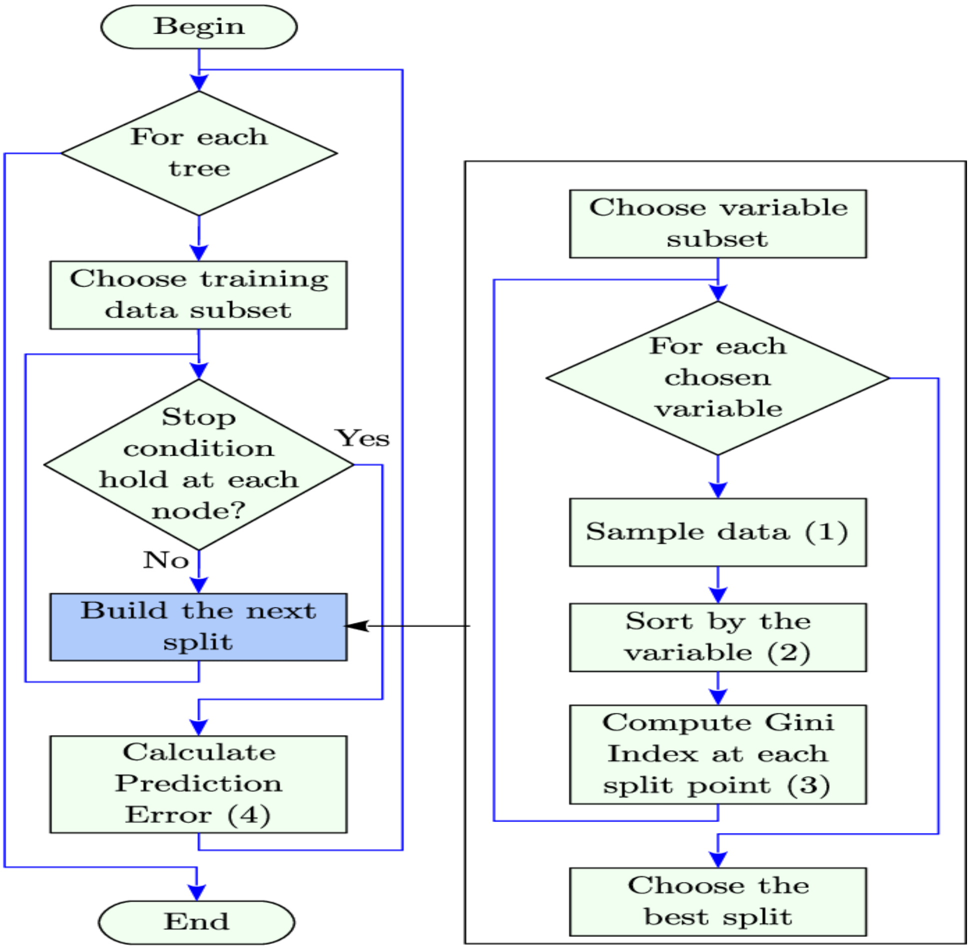

Figure 2: Flow chart of random forest algorithm

Evaluation Criteria

The model's performance was evaluated using performance metrics such as confusion matrix, accuracy, precision, recall, accuracy, and F1 score.

Confusion Matrix

A confusion matrix is an n*n matrix (where n is the number of labels) used to describe the performance of a classification model. It contains several true positives, true negatives, false positives, and false negatives. True positives are the number of cases where the model correctly predicted a positive outcome and true negatives are the number of cases where the model correctly predicted a negative outcome. False positives are the number of cases where the model predicted negative outcomes as positives, while false negatives are cases where the model predicted positive outcomes as negatives.

Accuracy

Accuracy is a performance metric used to measure the percentage of correctly classified observations. The mathematical expression for calculating machine accuracy is stated by Equation 2 (Guermoui *et al.*, 2018).

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Sample}} \qquad (2)$$

Precision

Precision measures the proportion of true positives among the positive predictions. It is obtained by dividing the number of true positives by the sum of true positives and false positives. The machine's precision is calculated by using Equation 3 (Guermoui *et al.*, 2018).

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positive (TP)} + \text{False Positives(FP)}} \qquad (3)$$

Recall

Recall, also called "true positive rate" or sensitivity," measures the proportion of true positives among the actual positives. It is calculated by dividing the number of true positives by the sum of true positives and false negatives.

F1-Score

The F1 score is a metric that combines precision and recall into a single core. Equation 4 is used for finding the F1 value (Chen *et al.*, 2011).

$$\text{F1-Score} = \text{Precision} + \text{Recall} \qquad (4)$$

Data Visualization

The data is made up of ten thousand (10000) target data sets, comprising one for zero power failure target output and one for power failure target output.

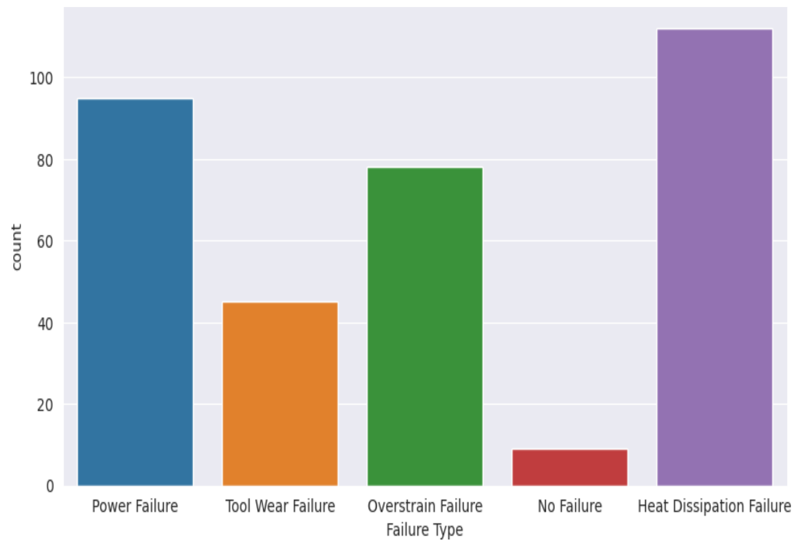

Fig. 3: Target value on causes of power failure

Figure 3 shows the expanded causes of power failure based on tool wear failure, overstrain failure, no failure, and heat dissipation failure.

RESULTS AND DISCUSSION

Performance Metrics of a Random Forest Model

A random forest model was developed in this study using five input parameters: air temperature, process temperature, rotational speed, torque, and tool wear. Table 1 shows the results of the performance metrics for the built random forest model.

Confusion Matrix Analysis

A confusion matrix was built using validation data ciproxifan seen data to evaluate how well the model has learned in classifying machine failure and no machine failure from the data set. The true label signifies the actual target of our data set, while the predicted label depicts the model target output. Summing up the rolls, there were 2899 no-machine failure targets and 101 machine failure targets. However, of the 2899 no machine failure targets, the model predicted 2884 no machine failures as true positives (TP) and 15 no machine failures as false positives (FP) better in performance as contained in Benmouiza and Cheknane, (2013). On the other hand, the developed model precisely predicted 72 of the machine failure targets as true negatives (TN) out of 101 machine failure targets and 29 as false negatives (FN).

Table1: Performance metrics of random forest model

Model	Accuracy	Precision	Recall	F1
Training	0.9934	0.8403	0.96153	0.8967
Testing	0.9853	0.7128	0.82758	0.7660

Figure 4 shows the possibility of machine failure with respect to the five input parameters. The possibility of machine failure increases as air temperature increases from 296K to 302K before decreasing sharply to 304K. For air temperatures above 304K, the possibility of machine failure is quite high. Whereas, an increase in process temperature has little effect until the process temperature exceeds 313K. Similarly, an increase in rotational speed increases the possibility of machine failure, specifically when it is greater than 2600 rpm leading to machine failure. Meanwhile, an increase in torque between 10 Nm and 40 Nm does not affect the possibility of machine failure except at 50 Nm and above, which increases the possibility of machine failure until 70 Nm, when the machine fails. An increase in tool wear from 0 to 150 minutes does not affect the possibility of machine failure; rather, the possibility of machine failure increases for tool wear between 200 and 250 minutes, with failure occurring at 250 and above the report of Torabi *et al.,* (2018).

CONCLUSION

Predictive maintenance models are designed to help determine the condition of machines to predict when maintenance is needed to avoid them breaking down completely. It is based on the collection, preprocessing, training, and intelligent use of data, which allows for safety compliance, preemptive corrective actions, and increased asset life. This study developed and evaluated a random forest machine learning model on generic algorithm for predicting milling machine failure using five input parameters.

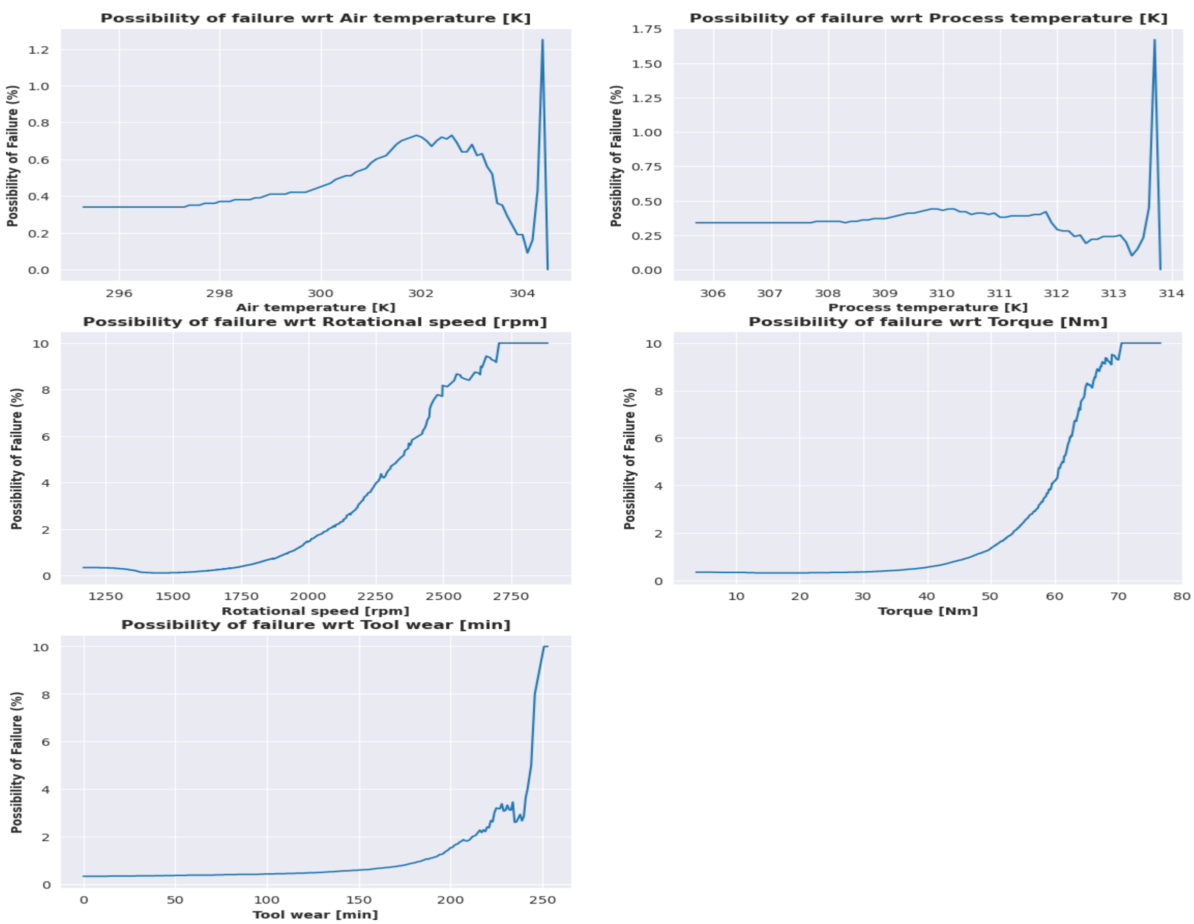

Fig. 4: Possibility of machine failure with input parameters

Deploying the model helps operators know when failure is likely to occur and what measures can be taken before the machine fails, including pre-emptive investigation, maintenance schedule adjustments, and sand repairs. The performance evaluation of the developed random forest model revealed that it predicts milling machine failure with high precision and accuracy, as evidenced by the performance metrics obtained during the model testing: accuracy, precision, recall, and F1-stated values.

The model helped determine the downtime and precise time for preventive maintenance to take place. This would in turn increase productivity and machine efficiency, as well as elongate its life span.

REFERENCES

Achour, L., Bouharkat, M., Assas, O., and Behar, O. (2017). Hybrid model for estimating monthly global solar radiation for the southern of Algeria :(case study: Tamanrasset, algeria). Energy, 135, 526–539.

Benmouiza, K., and Cheknane, A. (2013). Forecasting hourly global solar radiation using hy-bridk-means and non-linear autoregressive neural network models Energy ConversionandManagement,75, 561–569.

Chen, J.-L. and Li, G.-S. (2014). Evaluation of support vector machine for estimation of solar radiation from measured meteorological variables. Theoretical and applied climatology,115(3), 627–638.

Chen, J.-L., Li, G.-S., and Wu, S.-J. (2013). Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energy conversion and management, 75, 311–318.

Chen, J.-L., Liu, H.-B., Wu, W., and Xie, D.-T. (2011). Estimation of monthly solar radiation from measured temperatures using support vector machines–a case study. Renewable Energy, 36(1), 413–420.

Diez-Olivan, A., Pagan, J A., Sanz, R. and Sierra, B. (2017). Data-driven prognostics using a combination of constrained k-means clustering, fuzzy modeling and of-based score. Neurocomputing, 241, 97–107.

Gala,Y., Fern´andez, A´., D´ıaz, J. and Dorronsoro, J. R. (2016). Hybrid-machine learning Forecasting of solar radiation values. Neurocomputing, 176, 48–59.

Grinˇcov´a, A. and MArAsov´a, D. (2014). Experimental research and mathematical modelling an effective tools for assessing failure of conveyor belts. Eksploatacjai Niezawodno´s´c, 16(2).

Guermoui, M., Rabehi, A., Gairaa, K. and Benkaciali, S. (2018). Support vector regression methodology for estimating global solar radiation in Algeria. The European PhysicalJournalPlus, 133(1), 1–9.

Herath, H., Ariyathunge, S., and Karunasena, G. (2021). Development of a mathematical model to forecast solar radiation and validating results using machine learning techniques. Renewable energy, 36 (8), 213–225.

Jansen, F., Holenderski, M., Ozcelebi, T., Dam, P., and Tijsma, B. (2018). Predicting machine failures from industrial time series data. In 2018 5th International conference on control, decision and information technologies (CODIT), 1091-1096. doi:10.1109/CoDIT.2018.8394915

Paller,G., & E´l¨o,G. (2022). Towards a floating plastic waste early warning system. Sensor nets, 5(1), 45–50.

Park, Y., Cho, K. H., Park, J., Cha, S. M., and Kim, J. H. (2015). Development of an early-warning protocol for predicting chlorophyll-a concentration using machine learning models in fresh water and estuarine reservoirs, Korea. Science of the Total Environment, 502, 31–41.

Rahimikhoob, A. (2010). Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. Renewable energy, 35 (9), 2131–2135.

TienBui, D., Moayedi, H., G¨or, M., Jaafari, A. and Foong, L K (2019). Predicting slope stability failure through machine learning paradigms. ISPRS International Journal of Geo-Information,8(9), 395.

Torabi, M., Mosavi, A., Ozturk, `P., Varkonyi-Koczy, A., & Istvan, V. (2018). A hybrid machine learning approach for daily prediction of solar radiation. In International conference on global research and education, 9(2), 266–274.