Title Page:

Innovative Method of Predicting Maintenance for Industrial Machinery using Supervised Machine Learning in Comparison with Random Forest for Improved Accuracy

Madhav Amilineni[1], Nagaraj[2]

Madhav Amilineni[1],
Research Scholar,
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105
mahaboobbashas.sse@saveetha.com

Nagraj[2]
Research Guide, Corresponding Author,
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105
.sse@saveetha.com

Keywords: Condition Monitoring, Equipment, Fault Detection, Novel Supervised Machine Learning, Reduction, Random Forest, Sensor Fusion.

ABSTRACT

Aim: The goal of this study is to innovatively Predict Maintenance for Industrial Equipment by Comparing Supervised Machine Learning to Random ForestAccuracy. **Materials And Methods:** Using characteristics, the industrial equipment dataset is utilised to forecast maintenance of industrial machinery. The dataset contains equipment features that aid in maintenance prediction, and it is divided into two samples. The dataset is tested at 80% for g power with a 95% confidence level using the t-test analysis, and it is divided into two samples, group1 containing (n=10) images and group2 containing (n=10) images. **Results:** According to SPSS statistical analysis, supervised machine learning has an accuracy of 99% and a loss of 1%, which appears to be superior than the random forest's accuracy of 98% and a loss of 2%, with a statistical significance difference of $p=0.001$ ($p<0.05$). **Conclusion:** According to the findings, supervised machine learning greatly outperforms random forest in terms of accuracy when predicting the maintenance of industrial equipment.

Keywords: Condition Monitoring, random forest, Equipment, Fault Detection, Novel Supervised Machine Learning, Reduction, Sensor Fusion.

INTRODUCTION

The study's goal was to replace these losses and malfunctions with predictive maintenance using sophisticated machine learning techniques. Discover and predict problems in industrial machinery using Machine Learning (ML) approaches (Zhang et al. 2019). Manufacturing firms are under pressure to be more adaptable, reduce downtime and costs, and increase productivity. Responding to these pressures, data-driven manufacturing companies are leveraging the capabilities of artificial intelligence (AI), the industrial internet of things (IIoT), cloud computing technologies, and innovations in smart measurement and quality data management systems, resulting in greater visibility into their operations (Cakir, Guvenc, and Mistikoglu 2021). Inadequate maintenance practises can already reduce total productivity by 5 to 20%. Equipment failure prediction, anomaly detection, cost analysis, human error prediction, continuous monitoring, and feedback loop are some of the applications (Serradilla et al. 2022) (Zermane and Drardja 2022).

There are around 173 articles published in IEEE and 84 articles published in google scholar for the past 5 years. (Leukel, González, and Riekert 2021) have proposed a paper on systematic review of adopting the machine learning technology failure prediction in industrial maintenance. (Md et al. 2022) proposed a review on data driven quality prediction in the process with machine learning for industry 4.0. (Bakdi, Kristensen, and Stakkeland 2022) proposed paper on multiple instance learning with random forest for event logs analysis and predictive maintenance in ship

electric propulsion system. (Fernandes, Corchado, and Marreiros 2022) proposed a paper on machine learning techniques applied to medical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases with systematic literature review. (Bertolini et al. 2021) proposed a comprehensive literature review on machine learning for industrial applications.

The study gap highlighted that present research has a lack of precision in estimating and predicting industrial machinery maintenance. There are various disadvantages to supervised machine learning, which include Supervised learning necessitates a labeled dataset, which necessitates a significant amount of data with precisely labeled examples. Data labeling may be time-consuming and expensive, especially in complex settings. The current study aim was designed based on the aforementioned difficulties to anticipate the maintenance of industrial equipment using supervised machine learning.

MATERIALS AND METHODS

The suggested research is conducted in the Saveetha School of Engineering's Information Security Laboratory. There were a total two groups employed in the investigation. Supervised machine learning (Group 1=) and random forest (Group 2=). There are 10 samples in each category (Zenisek, Holzinger, and Affenzeller 2019).

The information is used to forecast the maintenance of industrial machines. The data was obtained from kaggle.com (Nagaraj et al. 2022). Each group receives 10 samples in total. The dataset is divided into testing and training data. Ten samples are gathered for the test data and ten for the training data. The dataset is partitioned, and the algorithm is then fitted to forecast the accuracy values, create, and test sets (Jiang et al. 2022). Clinccalc.com is used to compute the sample size, using G-power of 0.8, alpha and beta values of 0.05 and 0.2, and a confidence interval of 95%. In this work, the accuracy of supervised machine learning and random forest is examined (Ahmed, Jeon, and Piccialli 2022).

Supervised Machine Learning

Supervised machine learning is a subfield of AI and machine learning in which a model is taught to learn a mapping from input data to corresponding output labels. The term "supervised" implies that the algorithm is provided with a labelled dataset that includes both the input features and the proper output labels. The goal is to learn a mapping or function that can make accurate predictions or classifications based on new, previously unknown data.

Pseudocode

Step 1 : Import the dataset and the required packages.
Step 2 : Specify what X and Y are.
Step 3 : Establish training and testing sets for the data.
Step 4 : X train, x test, y test, & train test split(x,y,random state=50,test size=0.3)
Step 5 : Fit Model SVM (x train, y train)
Step 6 : Model svm.predict(x test) is used.
Step 7 : Model SVM.Score(x test, y test)
Step 8 : Display the Accuracy.

Random Forest

Random Forest is a machine learning ensemble approach that may be used for classification as well as regression applications. It is noted for its robustness and efficacy in a wide range of applications and is based on decision tree algorithms. Random Forest has evolved into a formidable machine learning tool, praised for its adaptability and predicted accuracy.

Pseudocode

Step 1 : Import the dataset and the required packages.
Step 2 : Specify what X and Y are.
Step 3 : Establish training and testing sets for the data.
Step 4 : X train, x test, y test, & train test split(x,y,random state=50,test size=0.3)
Step 5 : Fit Model RF (x train, y train)
Step 6 : Model rf.predict(x test) is used.
Step 7 : Model RF.Score(x test, y test)
Step 8 : Display the Accuracy.

A system with Windows OS and Hard disk capacity of 65 GB is used. Ram of 4GB and Language used id Python, either implemented in Jupyter (Anaconda). Processor used is Intel (R) Core (TM). Independent variables for equipment value. Dependent variables are improved accuracy values.

STATISTICAL ANALYSIS

The IBM SPSS version saas was utilized as a statistics programme for this investigation. The dependent variable is to increase accuracy metrics, while the independent variable is equipment value. In this investigation, a separate t-test analysis was performed (Pallathadka et al. 2022).

RESULTS

Table 1 represents the dataset values of supervised machine learning and Random forest.

Table 2 represents aggregate statistics of 10 samples that provide the accuracy mean and statistical loss of Supervised machine learning got 99% accuracy & 1.00% loss appears to be more when compared with Random Forest of 98% and 2% loss. The calculation of standard deviation and mean errors (standard error mean for supervised machine learning is .13944 and Random forest is .15770).

Table 2 shows Independent test analysis, it gives significance ($p<0.05$).Significance valued for Accuracy is 0.001. Standard deviation for Accuracy is. T-test equality for means include significance ($p<0.05$) and standard error means and mean difference. For each algorithm 20 samples are taken for supervised machine learning and Random forest.

Fig. 1 shows a bar Chart representing the mean accuracy of random forest and supervised machine learning methods is compared. The mean accuracy of supervised machine learning is higher than that of random forest, while the standard deviation of supervised machine learning is somewhat lower than that of random forest. Random forest approaches tend to be more accurate than Supervised machine learning.

DISCUSSION

In this investigation, we found that the Novel supervised machine learning was superior to the random forest in terms of precision and error difference ($p=0.001$, Independent Sample T Test). As opposed to random forest(mean accuracy = 99.00, mean loss = 1.00), supervised machine learning had better accuracy and a lower loss (mean accuracy = 98.00, mean loss = 2.00).

The similar findings of the related work found in the previous study are discussed. (Nagaraj et al. 2022) proposed a paper on ensemble machine learning(grid search and random forest) based on the enhanced medical expert recommendation system for diabetes mellitus prediction . (Serradilla et al. 2020) proposed a paper based on interpreting remaining useful life estimations combining explainable artificial intelligence and domain knowledge in industrial machinery. (Pallathadka et al. 2022) proposed a review of using artificial intelligence and machine learning in the food and agriculture industry. Dissimilar findings for related papers. (Calabrese et al. 2020) proposed SOPHIA - an event based IoT and machine learning architecture for predictive maintenance in industry 4.0. (Vivek and Kusuma 2023) proposed an effective quality machine learning to predict the equipment industry. (Das and Rangarajan 2020) conducted a survey based on performance monitoring and failure prediction of industrial equipment using artificial intelligence and machine learning methods.

We have comparable and contrasting conclusions from the preceding discussion, where we notice an increase in success rate and accuracy. The execution time of both methods varies. We may infer that enhanced supervised machine learning looks to have higher accuracy and performance than random forest approaches in all scenarios. Random forest approaches have the disadvantage of being computationally sluggish and taking a long time to train. In the future, we will overcome random forest approaches' constraints such as algorithm and training speed by introducing additional training data attributes. We can also test these algorithms on considerably bigger datasets.

CONCLUSION

This work used two classification approaches, supervised machine learning and random forest, to measure the quantity of chlorophyll in tea leaves. Random forest appears to perform worse than supervised machine learning. When it comes to estimating maintenance costs for industrial equipment, novel supervised machine learning looks to outperform random forest.

DECLARATION

Conflict of Interests

This manuscript has no conflicts of interest.

Authors Contribution

Data gathering, analysis, and article writing are all tasks that author PA is involved in. Conceptualization, data validation, and critical assessment of the text were all done by author KVK.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study:

1.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

Ahmed, Imran, Gwanggil Jeon, and Francesco Piccialli. 2022. “From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where.” *IEEE Transactions on Industrial Informatics* 18 (8): 5031–42.

Bakdi, Azzeddine, Nicolay Bjørlo Kristensen, and Morten Stakkeland. 2022. “Multiple Instance Learning With Random Forest for Event Logs Analysis and Predictive Maintenance in Ship Electric Propulsion System.” *IEEE Transactions on Industrial Informatics* 18 (11): 7718–28.

Bertolini, Massimo, Davide Mezzogori, Mattia Neroni, and Francesco Zammori. 2021. “Machine Learning for Industrial Applications: A Comprehensive Literature Review.” *Expert Systems with Applications* 175 (August): 114820.

Cakir, Mustafa, Mehmet Ali Guvenc, and Selcuk Mistikoglu. 2021. “The Experimental Application of Popular Machine Learning Algorithms on Predictive Maintenance and the Design of IIoT Based Condition Monitoring System.” *Computers & Industrial Engineering* 151 (January): 106948.

Calabrese, Matteo, Martin Cimmino, Francesca Fiume, Martina Manfrin, Luca Romeo, Silvia Ceccacci, Marina Paolanti, et al. 2020. “SOPHIA: An Event-Based IoT and Machine Learning Architecture for Predictive Maintenance in Industry 4.0.” *Information. An International Interdisciplinary Journal* 11 (4): 202.

Das, Mukunda K., and Krishnan Rangarajan. 2020. “Performance Monitoring and Failure Prediction of Industrial Equipments Using Artificial Intelligence and Machine Learning Methods: A Survey.” In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 595–602. IEEE.

Fernandes, Marta, Juan Manuel Corchado, and Goreti Marreiros. 2022. “Machine Learning Techniques Applied to Mechanical Fault Diagnosis and Fault Prognosis in the Context of Real Industrial Manufacturing Use-Cases: A Systematic Literature Review.” *Applied Intelligence (Dordrecht, Netherlands)* 52 (12): 14246–80.

Leukel, Joerg, Julian González, and Martin Riekert. 2021. “Adoption of Machine Learning Technology for Failure Prediction in Industrial Maintenance: A Systematic Review.” *Journal of Manufacturing Systems* 61 (October): 87–96.

Md, Abdul Quadir, Keshav Jha, Sabireen Haneef, Arun Kumar Sivaraman, and Kong Fah Tee. 2022. “A Review on Data-Driven Quality Prediction in the Production Process with Machine Learning for Industry 4.0.” *Processes* 10 (10): 1966.

Nagaraj, P., V. Muneeswaran, Sudar K. Muthamil, Vardhan Reddy A. Naga, G. Deshik, and Kumar Reddy C. Charan. 2022. “Ensemble Machine Learning (Grid Search & Random Forest) Based Enhanced Medical Expert Recommendation System for Diabetes Mellitus Prediction.” In *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 757–65. IEEE.

Pallathadka, Harikumar, Malik Jawarneh, F. Sammy, Vipul Garchar, Domenic T. Sanchez, and Mohd Naved. 2022. “A Review of Using Artificial Intelligence and Machine Learning in Food and Agriculture Industry.” In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2215–18. IEEE.

Serradilla, Oscar, Ekhi Zugasti, Carlos Cernuda, Andoitz Aranburu, Julian Ramirez de Okariz, and Urko Zurutuza. 2020. “Interpreting Remaining Useful Life Estimations Combining Explainable Artificial Intelligence and Domain Knowledge in Industrial Machinery.” In

2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–8. IEEE.
Serradilla, Oscar, Ekhi Zugasti, Jon Rodriguez, and Urko Zurutuza. 2022. "Deep Learning Models for Predictive Maintenance: A Survey, Comparison, Challenges and Prospects." *Applied Intelligence* 52 (10): 10934–64.
Vivek, B. V., and S. M. Kusuma. 2023. "An Effective Quality Machine Learning to Predict Industry Equipment Failures." In *2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 1–6. IEEE.
Zenisek, Jan, Florian Holzinger, and Michael Affenzeller. 2019. "Machine Learning Based Concept Drift Detection for Predictive Maintenance." *Computers & Industrial Engineering* 137 (November): 106031.
Zermane, Hanane, and Abbes Drardja. 2022. "Development of an Efficient Cement Production Monitoring System Based on the Improved Random Forest Algorithm." *International Journal of Advanced Manufacturing Technology* 120 (3): 1853–66.

TABLES AND FIGURES

Table 1: Accuracy values of Supervised machine learning and Random forest

S.No	SVM	Random forest
1	99.89	98.87
2	99.87	98.85
3	99.85	98.82
4	99.83	98.79
5	99.80	97.98
6	99.75	97.94
7	99.69	97.90
8	98.93	97.85
9	98.90	97.83
10	98.89	97.75

Table 2. Group statistics of accuracy for Novel Supervised machine learning and Random forest

	GROUPS	N	MEAN	STD.DEVIATION	STD.ERROR MEAN
ACCURACY	SVM	10	99.5400	.44096	.13944
	RANDOM FOREST	10	98.2580	.49868	.15770

Table 2. Independent Samples T-Test for Novel 1-D Convolutional Neural Network and Multiple Linear Regression

	Levene's Test for Equality of Variances		**T-Test for Equality of Mean**					**95%Confidence Interval of Difference**	
	F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**Lower**	**Upper**
Equal variances assumed	1.321	.266	6.090	18	.000	1.2820	.21051	.83974	1.72426
Equal variances assumed			6.090	17.734	.000	1.2820	.21051	.83927	1.72473

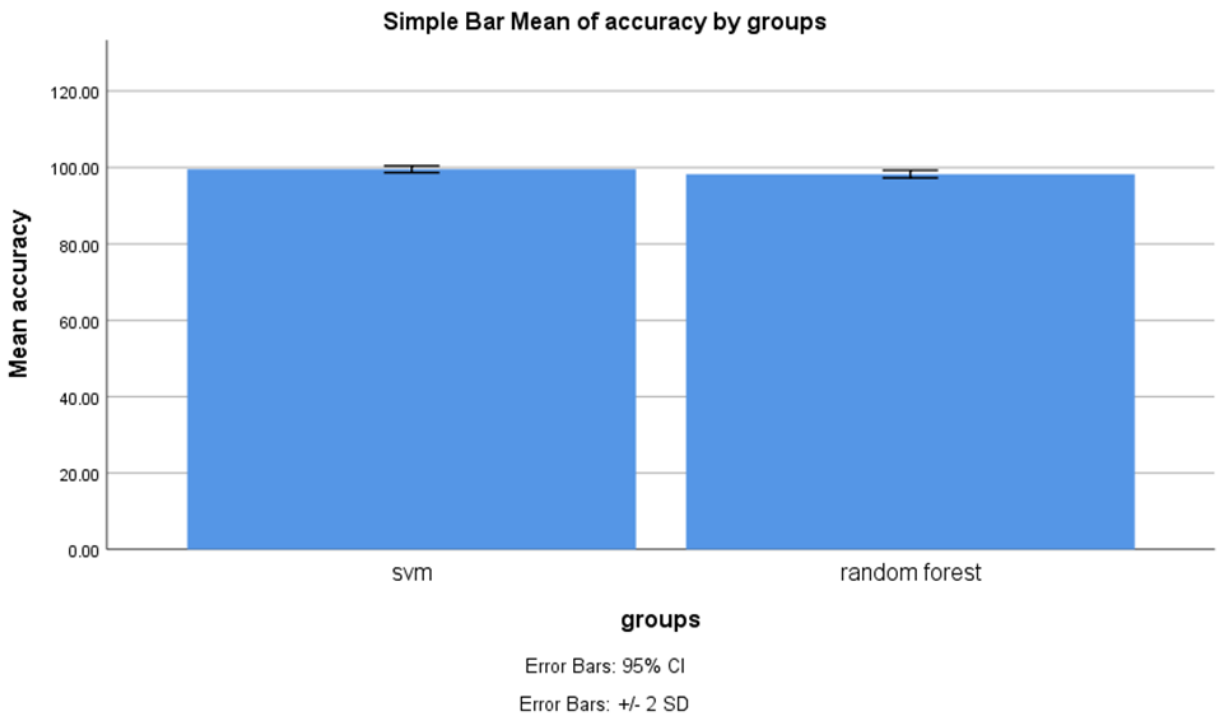

Fig.1. Bar Chart representing the comparison of Mean Accuracy of SVM and random forest algorithms. Mean accuracy of SVM is better than random forest and standard deviation is slightly better for SVM than random forest. X-Axis: SVM vs random forest, Y-Axis: Mean Accuracy of detection ± 2 SD.