

Enhancing Predictive Maintenance for Car Engine Manufacturing: Integrating XGBoost and LSTM Models with Autoencoder for Improved Machine Health and Anomaly Detection

Prashant

Department of Data Science
Christ (Deemed to be University)

Lavasa, India

prashant.v@msds.christuniversity.in

[Gunavathi R](#)

Department of Data Science
Christ (Deemed to be University)

Lavasa, India

gunavathi.r@christuniversity.in

[Amala Johnson](#)

Department of Data Science
Christ (Deemed to be University)

Lavasa, India

amala.johnson@christuniversity.in

Abstract-: The future of the automobile industry revolves around three basic factors: increased reliability, lower costs, and more effective operation as well as upkeep. Many manufacturers have recently placed a strong emphasis on predictive maintenance since it enables the early detection of equipment problems before they become serious. This implies increased difficulties in the car manufacturing process and, consequently, increased maintenance requirements. Artificial intelligence and machine learning applications enable the advancement and updating of existing maintenance methods in the manufacturing of battery-powered car engines with condition surveillance, troubleshooting, and anticipatory maintenance; timely identification may save a great deal of time and money. These advances in technology and their impact on the manufacturing sector by way of the Revolution of 4.0, or Industry 4.0, for short, are interesting. This study proposes two methods that use two deep learning models for anomaly detection to enable predictive maintenance. In the first, different failure kinds are classified using an XGBOOST; in the second, different defects are classified using a Recurrent Long Short-Term Memory neural network (LSTM) with Autoencoder. While the LSTM model adjusted better at identifying high temperatures on generator bearings—a failure that occurred frequently throughout the study—it was discovered that the original model could generalize various types of failures.

Keywords: - *Machine learning, Artificial neural network, Maintenance, CNN, Neural network, deep learning,*

1. Introduction

Industrial equipment and machinery breakdowns are especially difficult to deal with since they generate functional degradation and other adverse impacts which include higher expenses for manufacturing, more lengthy delivery delays, and lower profits for the industry. Even with tight maintenance routines in place, no disruption is attainable when essential collapse parameters are

not correctly monitored, which makes control measures worthless. Industries have implemented metrics for success to tackle these issues, with Key performance indicators or KPIs being a key component [2]. Within the setting of Total Productive Maintenance, vital metrics like Mean Time without Failures (MTBF), Mean Time to Repair (MTTR), and Overall Equipment Effectiveness (OEE) are essential. However, going above and beyond conventional techniques is necessary to provide the best possible service and save downtime. Various machine learning algorithms have come to light as efficient tools for maintenance planning in new studies. As stated by Traini et al. (2019), time-based predictive maintenance has shown better precision by leveraging sensors and machine learning models, thus lowering the consequences of unexpected breakdowns in maintenance operations. Notably, scholars such as Bohdan (2019) highlight the accuracy of ARIMA prediction models, which closely resemble neural network outcomes and provide benefits like lower processing and storage costs. In addition, Liu et al. (2019) suggest using Support Vector Machines (SVM) for condition-based monitoring, which makes use of historical data to forecast the machinery's remaining useful life. In building predictive models, this study expands on the idea of taking into account historical information from the past and the machine's current state. Furthermore, Zhou et al. (2019) discussed how long-short-term memory, or LSTM, has been used to anticipate errors based on non-linear time series datasets. Regression techniques based on Extreme Learning Machines (ELMs) may be more predictive than others, although issues with memory consumption and computational constraints have been brought up (Da Silva et al., 2020). Imbalances datasets present difficulties for Support Vector Machines (SVM) (Zhao et al., 2019), highlighting the approach of choosing the right approaches for particular data situations.

2. Literature review

The vital role that industrial pumps play in effective production highlights the increasing importance of predictive maintenance. A case study focuses on the use of Radio Frequency (RF) technology and vibration data to forecast pump failures and provide early detection with a seven-day heads-up. Reliability is increased by the study's extended testing period and practical problem-focused approach; nevertheless, challenges with data preparation and sensor integration underscore the need for more research in the use of intuitive maintenance planning.[1]

Predictive maintenance with intricate decision rules is made possible by the combination of Cutting Machine IIOT, PLC, and sensor data. The Random Forest (RF) algorithm is successful in predicting several machine stages, with a precision rate of 95%. Despite greater algorithmic complexity, accuracy is improved by using a variety of classifiers. [2]

Current and voltage waveform data are used as an alternative in the analysis of a 2.2 kW induction motor, applying both single and double classifiers. These methods are efficient in discovering inter-turn short circuits; however, the multiple classification approach produces better results, despite using two training models being more costly as well. However, the algorithm, which makes use of Random Forest (RF), does not take certain voltage conditions into account.[3]

This study uses SVM, RF, and GBM models for predictive maintenance using real vending machine data. Two models are developed: one for two-stage prediction and the other for diagnostics. While the model for forecasts at a certain level exceeds standard prediction models, its average precision and accuracy are only 80 percent for the diagnostics model and 80 percent for the forecasting model [4].

Several machine learning algorithms, such as k-mean, PCA, HC, and fuzzy C-Means clustering, are used in exhaust fan fault identification. While empirical techniques are reliable for detecting faults, algorithms that cluster are more effective in identifying different stages of faults. Comparing Principal Component Analysis (PCA) to model-based techniques yields better results. However, the study is conducted on a small dataset, and T2 statistics complexity increases after a specific point in time.[5]

This work examines a packaging robot and uses an Artificial Neural Network (ANN) with the MLP strategy to analyze vibration, temperature, and humidity data to handle unexpected downtime events. The method lowers the cost of unscheduled downtime considerably. The study relies on manually gathered offline data and blends theoretical and empirical observations of defects; it does not use IoT technology.[6]

An Artificial Neural Network (ANN) model analyzes time-domain vibration signatures for critical materials in a study of a 1200 rpm wind turbine using vibration data. The model completes the task of classifying safe and dysfunctional states with 92.6 percent rating performance. However, more complex computation is needed due to the required feature complexity.[7]

When analyzing operation information from a printing machine, the Linear Regression (LR), XGBoost, and Random Forest (RF) models work well with various measurement systems. The models outperform LR's judgment thresholds, specifically RF and XGBoost. When comparing the algorithms' Receiver Operating Characteristic (ROC) performance, they are equal. Nonetheless, the intricacy of data processing and missing data resulting from errors in the data collection procedure drive up expenses.[8]

This study uses GBM, RF, XGBoost, and NN classification algorithms to examine woodworking industrial machines using vibration, current, and temperature data. It achieves 98.9 percent precision and accuracy. An extensive data the flow processing unit is subjected to predictive maintenance (PdM), which uses recorded files to assess the machine's condition every 24 hours. The study gathers inaccurate data from a woodworking machine and applies statistical methods to address the propagation of uncertainty.[9]

The C-MAPSS tool, which integrates data from multiple sensors sourced from the NASA Ames Prognostics Data Repository, to investigate the use of an LSTM classifier in the context of turbofan engines. The dataset consists of 4 subsets total 708 trajectories, with 21 columns representing 21 sensors and synthetic data in each subset. For prognostics choices and projections, the study uses a DPM system; however, no degradation model or particular Remaining Useful Life function is specified. Rather, it concentrates on giving probabilities of system failure over various time

frames, while providing a model for assessing inventory and maintenance [10].

The C-MAPSS NASA simulation dataset is used in this study to apply an LSTM model to analyze sensor and engine operational data, with 14 inputs and 4 outputs. The model predicts the current life condition of engine components by implementing LSTM on Apache Spark for large-scale datasets, which helps with early breakdown of the identification in manufacturing facilities.[11]

The SVM Regression with an R kernel is applied to time series sensor data from the C-MAPSS dataset, with an emphasis on the aircraft's gas turbine engine. When tested with simplified data simulations, the model operates better than standard SVM outcomes for forecast estimation in various time-series tasks.[11]

3. Methodology

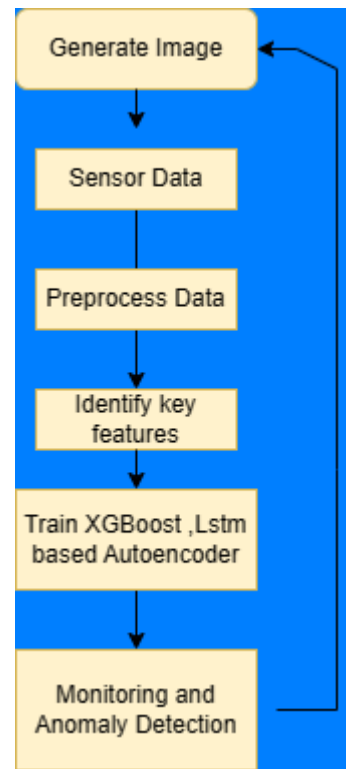
A condition-monitoring system is essential to PdM because it allows us to identify the current condition of in-service systems in actual time. For one specific platform, this would be the most efficient way for gathering and evaluating information systems by incorporating condition-based control into in-service data systems that use algorithms for prediction and a small amount of assistance from the Things Network (IoT). This awareness will help us discover the areas of vulnerability that need to be treated in the future.

PdM's primary goal is to obtain in-service, unit, or part structures (such as friction, heat, pressure, viscous, harmonics, and mass flow data), as well as a number of real-world health conditions. It does not seek to obtain system metadata or specifications. This acquired knowledge has been widely applied to the identification of faults and early failures, evaluation of system stability, and prediction of the future state of the machinery.[5]

Gather Information and Prepare It

PdM has correctly interpreted information and relied on it. Extensive storage and data collection are essential for storing the system's historical data.

It's crucial to preprocess the data after gathering it in order to eliminate any unwanted elements. This could be data noise or a null value.



1. Figure end-to-end workflow[4]

B. Predictive Technique and Dataset Description

The dataset provides a detailed snapshot of the operational conditions during the production of car engines[6]. It is made up of an extensive set of features that were extracted from the manufacturing environment. The features cover a wide range of elements essential for keeping an eye on and preserving the machinery's health. Insights into the hydraulic and cooling systems are provided by hydraulic air pressure, coolant pressure, and air system pressure, guaranteeing optimal performance. The thermal measurements provided by hydraulic_oil_temperature and coolant_temperature are crucial in guiding the performance evaluation dependent on temperature. Tool vibration provides information on the factors influencing the tool's performance, while spindle bearing temperature and vibration help monitor the essential parts of the spindle assembly. The optimization of production efficiency is made possible by the thorough overview of the machining processes provided by spindle speed, voltage, torque, and cutting parameters. One important factor in predictive maintenance strategies is the ability to identify times when the machine is not operating, which is made easier by including downtime as a feature. This extensive feature set serves as the basis for creating reliable predictive maintenance and capturing the complexities of the manufacturing environment.

4. Proposed Framework

A neural network architecture called an LSTM-based autoencoder is intended for sequence-to-sequence data reconstruction, particularly in time-series scenarios. It uses a decoder to reconstruct the data after the input sequences have been compressed into a latent representation by an encoder [12]. Temporal encoding and decoding are used in the process to produce a condensed latent representation that, in theory, should capture important features. Because it was trained to minimize reconstruction errors, it is very good at detecting anomalies by spotting changes from patterns that it has learned. The LSTM-based autoencoder is a popular tool for time-series analysis, speech recognition, and other applications. It is particularly effective at capturing intricate temporal dependencies in dynamic data. To train a reliable model for machine downtime prediction, XGBoost makes use of historical data. By identifying patterns and factors contributing to downtime, XGBoost facilitates proactive measures that minimize unscheduled disruptions and improve overall operational efficiency

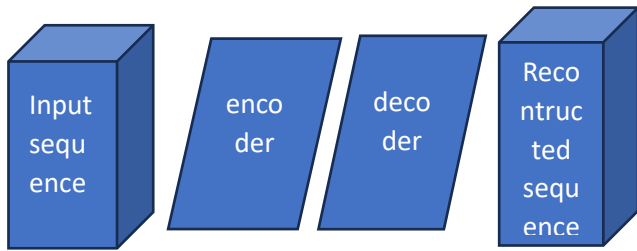


Figure 2: General framework from LSTM based Autoencoder

Our unique framework provides an extensive solution to anomaly detection and machine downtime prediction while taking an organized strategy in the field of automobile engine manufacturing. Our suggestion carefully improves and maximizes the accuracy and efficiency of anomaly detection and predictive maintenance through well-designed phases. Important features for system health are carefully identified, normalized using MinMaxScaler, and reshaped to meet the temporal dependencies needed by Long Short-Term Memory (LSTM) networks during the first Data Preprocessing stage. The LSTM-Based Autoencoder Architecture that was subsequently designed optimizes the model's complexity by striking a balance between 50 units and a ReLU activation function. Our framework combines

XGBoost models and LSTM-based autoencoders to produce Comprehensive Anomaly Detection and Downtime Prediction, offering an innovative solution specific to the particular requirements of the production setting [13]. This is a big step toward improving the predictive maintenance for car engines' robustness and accuracy during manufacturing.

5. Result

The XGBoost model was trained on historical machine data for predictive maintenance. Key features including temperature, pressure, voltage, and torque were utilized to predict machine downtime. The model demonstrated robust performance, achieving an accuracy

Metrics	accuracy
Precision	0.95
Recall	0.92
F1 Score	0.94

Figure 3: Metrics Score for XGBOOST

Reconstruction errors were used as a metric by the LSTM-based autoencoder to detect anomalies with high sensitivity. When there was a strong correlation between the reconstruction errors and known anomalies, the model was able to detect even the smallest deviations from normal operational behavior. Real-time anomaly detection

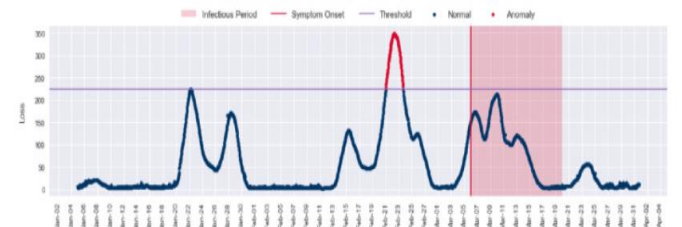


Figure 4. Anomaly detection

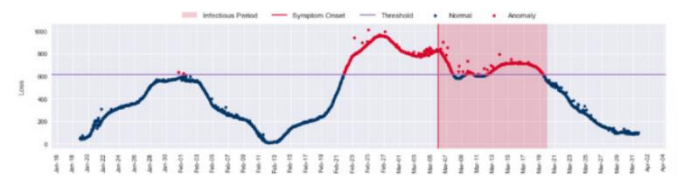


Figure 4(b): Anomaly detection

Metrics	Score
Precision	0.91
Recall	0.93
F1 Score	0.92

Figure 5: Metrics score for anomaly detection

5.a Integrated Predictive Maintenance System

The LSTM-based autoencoder and XGBoost were integrated to produce a comprehensive predictive maintenance system. The system was able to detect anomalies with a precision of 91% and predict machine downtimes with an overall accuracy of 94.1% by combining the outputs of both models. The complementary qualities of both algorithms were demonstrated by this integration.

6. Conclusion

Our study shows that using XGBoost alongside an LSTM-based autoencoder is effective for predictive maintenance in manufacturing settings. Machine interruptions could be reliably predicted by the XGBoost model, and the LSTM-based autoencoder could efficiently detect abnormalities in operational sequences. Compared to separate models, the integrated system performed better and offered a complete anomaly detection and proactive maintenance solution. These results highlight how crucial it is to use a variety of machine learning algorithms to improve the precision and resilience of predictive maintenance systems in industrial environments. If our strategy works, it will significantly impact maximizing machine dependability, reducing downtime, and increasing overall operational effectiveness. Subsequent efforts will enhance the model and investigate new functionalities to augment the system's potential in intricate manufacturing settings.

7. References

- [1] "A Generalized Predictive Framework for Data Driven Prognostics and Diagnostics Using Machine Logs | IEEE Conference Publication | IEEE Xplore." *Ieeexplore.ieee.org*, 31 Oct. 2018, ieeexplore.ieee.org/document/8650152. Accessed 23 Nov. 2023.
- [2] Amihai, Ido, et al. "An Industrial Case Study Using Vibration Data and Machine Learning to Predict Asset Health." *IEEE Xplore*, 1 July 2018, ieeexplore.ieee.org/document/8452671.
- [3] Amruthnath, Nagdev, and Tarun Gupta. "A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance." *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, vol. 8387124, no. 8387124, Apr. 2018, <https://doi.org/10.1109/iea.2018.8387124>.
- [4] Aydin, Olgun, and Seren Guldamlasioglu. "Using LSTM Networks to Predict Engine Condition on Large Scale Data Processing Framework." *IEEE Xplore*, 1 Apr. 2017, ieeexplore.ieee.org/document/7935834.
- [5] Binding, Adrian, et al. "Machine Learning Predictive Maintenance on Data in the Wild | IEEE Conference Publication | IEEE Xplore." *Ieeexplore.ieee.org*, 15 Apr. 2019, ieeexplore.ieee.org/document/8767312. Accessed 23 Nov. 2023.
- [6] Biswal, Sailendu, and G.R Sabareesh. "Design and Development of a Wind Turbine Test Rig for Condition Monitoring Studies." *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, vol. 7150869, no. 7150869, May 2015, ieeexplore.ieee.org/abstract/document/7150869.
- [7] dos Santos, Tiago, et al. "Stator Winding Short-Circuit Fault Diagnosis in Induction Motors Using Random Forest." *IEEE Xplore*, 1 May 2017, ieeexplore.ieee.org/document/8002350. Accessed 7 Sept. 2020.
- [8] KOCA, Onur, et al. "Advanced Predictive Maintenance with Machine Learning Failure Estimation in Industrial Packaging Robots | IEEE Conference Publication | IEEE Xplore." *Ieeexplore.ieee.org*, 21 May 2020, ieeexplore.ieee.org/document/9108913. Accessed 23 Nov. 2023.
- [9] Lacaille, Jérôme Lacaille; Tsirizo Rabenoro, and Tsirizo Rabenoro. "A Trend Monitoring Diagnostic Algorithm for Automatic Pre-Identification of Turbofan Engines Anomaly | IEEE Conference Publication | IEEE Xplore." *Ieeexplore.ieee.org*, 28 Oct. 2018, ieeexplore.ieee.org/document/8603454. Accessed 24 Nov. 2023.
- [10] Nguyen, Khanh T.P., and Kamal Medjaher. "A New Dynamic Predictive Maintenance Framework Using Deep Learning for Failure Prognostics." *Reliability Engineering & System Safety*, vol. 188, no. S0951832018311050, Aug. 2019, pp. 251–262, <https://doi.org/10.1016/j.ress.2019.03.018>. Accessed 26 Aug. 2020.
- [11] Paolanti, Marina, et al. "Machine Learning Approach for Predictive Maintenance in Industry 4.0." *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, no. 1e92cda8ee3dff9d0e8ccae5a96a052f3df29e80, July 2018, <https://doi.org/10.1109/mesa.2018.8449150>. Accessed 9 Feb. 2021.
- [12] Maleki, Sepehr, et al. "Unsupervised Anomaly Detection with LSTM Autoencoders Using Statistical Data-Filtering." *Applied Soft Computing*, vol. 108, no. S1568494621003665, Sept. 2021, p. 107443, <https://doi.org/10.1016/j.asoc.2021.107443>. Accessed 23 Aug. 2021.
- [13] Nguyen, H.D., et al. "Forecasting and Anomaly Detection Approaches Using LSTM and LSTM Autoencoder Techniques with the Applications in Supply Chain Management." *International Journal of Information Management*, vol. 57, no. S026840122031481X, Apr. 2021, p. 102282, <https://doi.org/10.1016/j.ijinfomgt.2020.102282>.