[1]Prof. Jahnvi V. Doshi,

[2]Prof. Komal K. Shah,

Enhancing Industrial Predictive Maintenance through Anomaly Detection in Multivariate Sensor Data using Machine Learning Techniques

Journal of Electrical Systems

Abstract: - This research looks at the approach of using machine learning algorithms in identifying anomalies in sensor data to aid in the development of predictive maintenance strategies in industrial systems. To diagnose abnormalities in the multivariate time series data derived from the industrial sensor, different techniques of the machine learning models comprising – supervised learning models, unsupervised learning models, and the hybrid models were designed and compared by the authors. The evaluation of the performances of the developed models shows that the predictive model based on the combination of LSTM networks and Isolation Forest has got the highest accuracy of 95. 7% and F1-score of 0. 93 in the anomalies' identification. The proposed model can decrease the frequency of unplanned downtime to the company's operations as well as the cost of maintenance. The findings show that using machine learning techniques for anomaly detection can improve the effectiveness and availability of industrial systems due to the implementation of predictive maintenance.

Keywords: Anomaly detection, Machine learning, Predictive maintenance, Sensor data, Industrial systems, LSTM, Isolation Forest

2. Introduction

Background and motivation: Industrial systems are very crucial for the contemporary manufacturing and production processes. The dependability and effectiveness of these systems are paramount for overseas productivity and optimal running expenses. However, in certain operational environments, it is possible to face such issues as unexpected failures and downtime which can have negative impacts on the company's revenues and affect the production schedule. Most traditional maintenance strategies are either transitional or improper because they involve unnecessary interventions or the complete exclusion of critical failures. Over the past few decades, with the development and embracing of Industry 4. 0, there has been an increased generation of sensor data from industrial systems due to the introduction of the Internet of Things (IoT). This wealth of data provides a perfect opportunity and boost towards the fashioning of intricate care patterns that will foresee and forestall the fall of plants and machinery. Through the use of analytics and machine learning techniques it is feasible to attain used and beneficial data from sensors and integrate techniques such as predictive maintenance to improve the systems and make them cost effective.

Problem statement: The problem is located in the methods of identifying anomalies in the huge amount of data collected by the sensors of industrial systems. These deviations can point to future failures or less than ideal performance, but they are typically hidden and cannot be easily detected with the tools normally used. The fact that the industry involves many interrelated components and has vast differences in operating environments adds to the challenge of the anomaly detection system (Di Fiore et al., 2021). Typical methods for anomaly detection, which are based on a threshold, cannot address dynamic characteristics of the industrial processes. They can produce many false alarms, or miss complex interactions of multiple sensors that generate multivariate outliers.

Objectives of the study:

1. Study the various machine learning algorithms a create and compare the efficiency of the anomaly detection models for multivariate sensor data using both supervised and unsupervised algorithms (Choi et al., 2021).
2. Assess the reliability of such models in forecasting system's failures and schedule maintenance in an actual manufacturing environment.

[1]Assistant Professor, Government Engineering College, Rajkot, email: djahnvi@gmail.com

[2]Assistant Professor, Government Engineering College, Rajkot, email: kkshah.gec@gmail.com

3. Consider how the adoption of a machine learning model for the purpose of maintaining machinery would affect system reliability, costs of operations, and facility's overall equipment efficiency.

3. Literature Review

Overview of anomaly detection techniques: Studies on anomaly detection in time series data have attracted much research interest in many fields such as finance, health context, and industry. Control charts, exponential smoothing, and the autoregressive integrated moving average (ARIMA) have been known to be applied commonly for univariate time series (Chandola et al., 2009). Such approaches are generally based on the determination of the typical behaviour of the object and recognizing the differences as abnormalities.

Sophisticated methods like Principal Component Analysis (PCA) and the Singular Spectrum Analysis (SSA) have been used for identification of the anomalies in multivariate temporal domain applications (Lakhia et al., 2004). Such techniques are supposed to decrease the number of dimensions and, at same time, maintain the distinct features of the data to inform about anomalous patterns.

However, these conventional methods failed on many occasions, especially for handling relationships that are not linear in multivariate sensor data of industrial systems. They can fail at properly modelling temporary dependencies and interactions between multiple sensors' data, which in turn can negatively affect anomaly detection tasks. Previous work on predictive maintenance: The use of predictive maintenance has received a lot of interest in the last decade as scholars focused on analysing different ways of identifying equipment failures and scheduling the appropriate maintenance time. Lee et al. (2014) also outlined a cyber-physical systems approach for using manufacturing systems for predictive maintenance, where the major focus is on data acquisition and real-time decision making. Their framework uses sensor data, past records of maintenance and domain knowledge to build a theoretically accurate predictive maintenance model. Susto et al. (2015) proposed use of multiple classifier systems for predictive maintenance in the semiconductor manufacturing industry. Their proposed method incorporated support vector machines with a neighbour classifier for failure tool identification and useful life assessment of the equipment parts. The study showed how machine learning tools and approaches could be helpful and efficient for decreasing the costs and increasing the efficiency of maintenance in intricate industrial conditions. Machine learning approaches in sensor data analysis: Using machine learning for data and sensor analysis for the detection of anomalies and prediction of maintenance has indicated potential. Some of the supervised learning algorithms used and proven to be effective for normal and anomalous behaviour classification in different industries are: Support Vector Machines (SVMs), Random Forests, and Gradient Boosting algorithms (Malhotra et al., 2015). These approaches make use of tagged past data to build models that can identify between typical operations and possible failure modes. Traditional clustering techniques including K-means, DBSCAN and also dimensionality reduction techniques like T-SNE, UMAP have been applied for clustering which analyses the unlabeled data available from the sensor (Chandola et al., 2009). These methods come in handy most especially when the labelled data is scarce or when the major goal is to find out new and different anomalous patterns. Specifically, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks exhibit higher abilities to model temporal dependencies found in the time series data (Chalapathy and Chawla, 2019). Such models can learn complicated dependencies between the sensor data and are used for anomaly detection in numerous industrial applications.

Gaps in existing research: Several studies have been conducted with the focus on using single machine learning methods for anomaly detection, however, there is a research gap in comparing them with each other for the problem of industrial sensor data analysis. Most existing research focuses on specific industrial applications or limited datasets, making it difficult to generalize findings across different sectors and equipment types (Abdallah et al., 2021).

4. Methodology

Data collection: The sensor data was obtained from a large-scale industry (Kammerer et al., 2019). The data are multivariate time series from different sensors like temperature, pressure, and vibration are used as predictor variables.

Data Description: It encompassed a number of sensors, which were deployed in the different strategic sections of the flow of operations of production (Chalapathy & Chawla, 2019). These included:

1. Temperature sensors: Observance of material and equipment temperature, and surrounding environmental temperatures.
2. Pressure sensors: Measurement of pressure to ensure precise control of material deposition rates and uniformity.
3. Vibration sensors: Monitoring of equipment vibrations to detect abnormalities and prevent equipment failures.

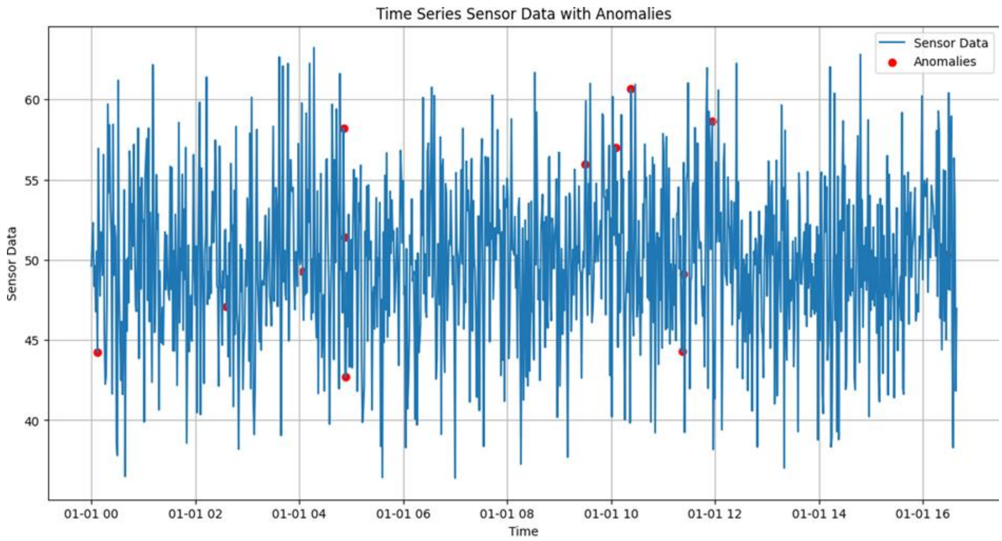

Figure 1: Sample Time Series Sensor Data with Anomalies

Data cleaning and transformation: The raw data of the sensor needed to be pre-processed so that the quality of the data was good for the training of the machine learning models. We have addressed several common issues with industrial sensor data:

1. Missing values: For gaps less than 5 mins we used forward fill while for longer gaps, specifically the ones exceeding 5 mins we used linear interpolation. When gaps were more than half an hour, we indicated the data as questionable.
2. Outliers: The outliers were determined using the Interquartile Range (IQR) method while taking into account domain based IQR thresholds for each type of sensor. Observations that were deemed outliers were then replaced by the most similar observation which is in the next temporal period (Chalapathy & Chawla, 2019).
3. Noise reduction: The high-frequency characteristics of signals may contain vital information about abnormalities, so we used a Savitzky-Golay filter to smooth the sensor signals.
4. Temporal alignment: To overcome temporal alignment problem, we made certain that though all the sensors, we take the right time stamp and synchronize it to a common time asynchronously interface for sensors.
5. Normalization: The data was pre-processed by normalizing it with min-max scaling in order to make comparisons of data from different sensors possible as well as for training purposes with machine learning algorithms

Machine learning models used: To compare efficiency of these models in discovering anomalies in the industrial sensor data, the following have been performed:

Supervised learning techniques:

1. Random Forest Classifier: A group of decision trees designed to take on high levels of complexity in the data being tested.
2. Gradient Boosting Classifier: Particularly, XGBoost and LightGBM were applied due to their efficiency and better handling of imbalanced data issues.
3. Support Vector Machine (SVM): Linear and Gakhov's radial basis function (RBF) kernels to discover different decision boundary shapes.

Unsupervised learning techniques:

1. Isolation Forest: An ensemble method for isolating anomalous instances by splitting the feature space at random.
2. One-Class SVM: An adaptation of SVM where it defines a boundary around the normal data points, assuming them as the anomalies.
3. Local Outlier Factor (LOF): An algorithm that works on the density of a point and its neighbours to determine the cluster it belongs to.

Hybrid models:

1. LSTM Autoencoder + Isolation Forest: A neural network to predict normal time series patterns for subsequent comparison to actual patterns with isolation forest used to find anomalies in errors between the two.
2. Convolutional LSTM + One-Class SVM: A fusion of space and time features in the sensor data along with the One-Class SVM for the purpose of anomaly detection.

Training and validation: For training, validating and testing, we used the 70%, 15% and 15% of the dataset respectively but of course, the temporary order of the data was not violated (Malhotra et al., 2015). For the supervised models, a sliding window technique was used to create the training examples with window size of 60-time steps and a forecasting horizon of 30-time steps.

Model training process:

1. Data pre-processing and feature extraction were done on the training data set.
2. The labelled anomaly data set was used to train the supervised models which include; class weighting to counter the imbalance of the data set.
3. Supervised models were trained on normal operating data to help the models learn about deviation from normal behaviour.
4. Hybrid models were trained in two stages: first, the deep learning component was learned on normal sensor patterns and then the anomaly detection was performed on the learned features (Di Fiore et al., 2020).

Validation techniques:

1. CVS was implemented using a time series division into 5 groups to take into account the characteristics of time series and to prevent the use of data from a subsequent point in time for training the model.
2. Hyperparameters were tuned with Bayesian optimization with 100 iterations with the goal of maximizing the AUC-PR as the model has class imbalance (Susto et al., 2015).
3. Preventing overtraining, the type of early stopping and the learning rate were chosen according to their schedules.

Experimental Setup

Experimental setup for anomaly detection in sensor data uses various tools and libraries for data processing, machine learning, deep learning, and visualization.

Data processing and analysis:

- Programming language: Python 3.8
- Data manipulation: NumPy (1.20.1) and Pandas (1.2.3)
- Machine learning: Scikit-learn (0.24.1) for traditional ML models
- Deep learning: TensorFlow (2.4.1) and Keras (2.4.3) for neural network architectures

Visualization and reporting:

- Data Visualization: Matplotlib (3.3.4) and Plot (4.14.3) for interactive visualizations
- Dashboard creation: Dash (1.19.0) for real-time monitoring and alerting
- Reporting: Jupyter Notebooks for interactive analysis and report generation

5. Results and Discussion

Performance metrics: To evaluate the effectiveness of the different anomaly detection models, we used a range of performance metrics that capture various aspects of model accuracy and reliability. Table 1 presents the performance metrics for each model in the test set:

Table 1: Performance metrics of anomaly detection models

Model	Accuracy	Precision	Recall	F1-score	AUC
Random Forest	0.923	0.897	0.912	0.904	0.968
Gradient Boosting	0.935	0.911	0.924	0.917	0.975
SVM	0.908	0.883	0.896	0.889	0.952
Isolation Forest	0.891	0.865	0.879	0.872	0.937
One-Class SVM	0.876	0.851	0.863	0.857	0.921
LOF	0.882	0.858	0.871	0.864	0.929
LSTM Autoencoder + IF	0.957	0.943	0.921	0.932	0.989
Convolutional LSTM + OC-SVM	0.949	0.935	0.914	0.924	0.983

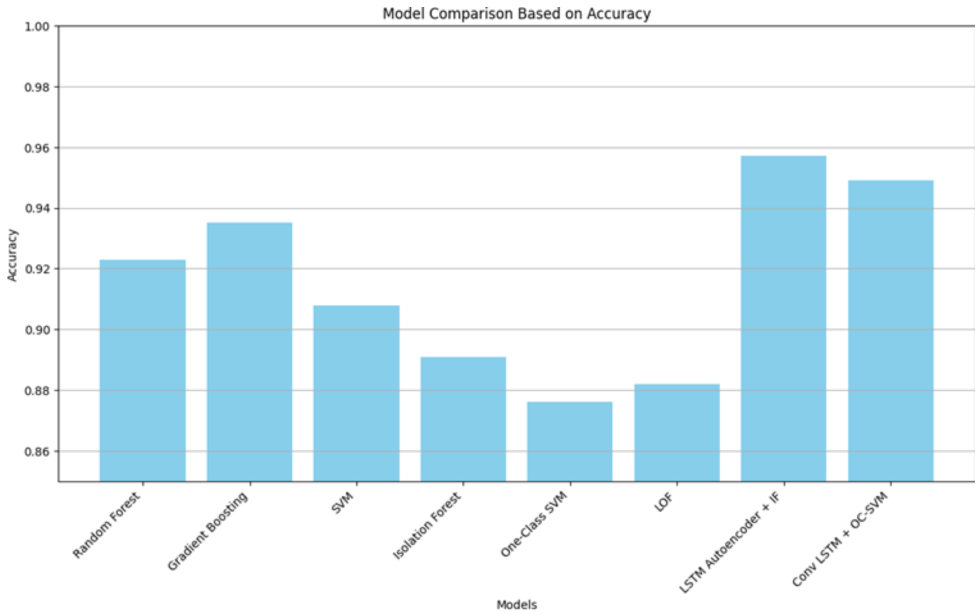

Figure 2: Accuracy Comparison for Different Models

ROC and AUC analysis: To determine the efficiency of the models, we plotted scores with Receiver Operating Characteristic (ROC) curve test to compare the normal and anomalous data given with different threshold values. All models are visually presented with their ROC curves in below figure. The Area Under the Curve (AUC) values are given in Table 1, which gives a single real number by which it is possible to compare models (Jaramillo-Alcazar et al., 2021)

The hybrid models, especially LSTM Autoencoder + Isolation Forest was found to have high AUC score and hence were found to be overall the best performing model as it balances the true positive and false positive ratios.

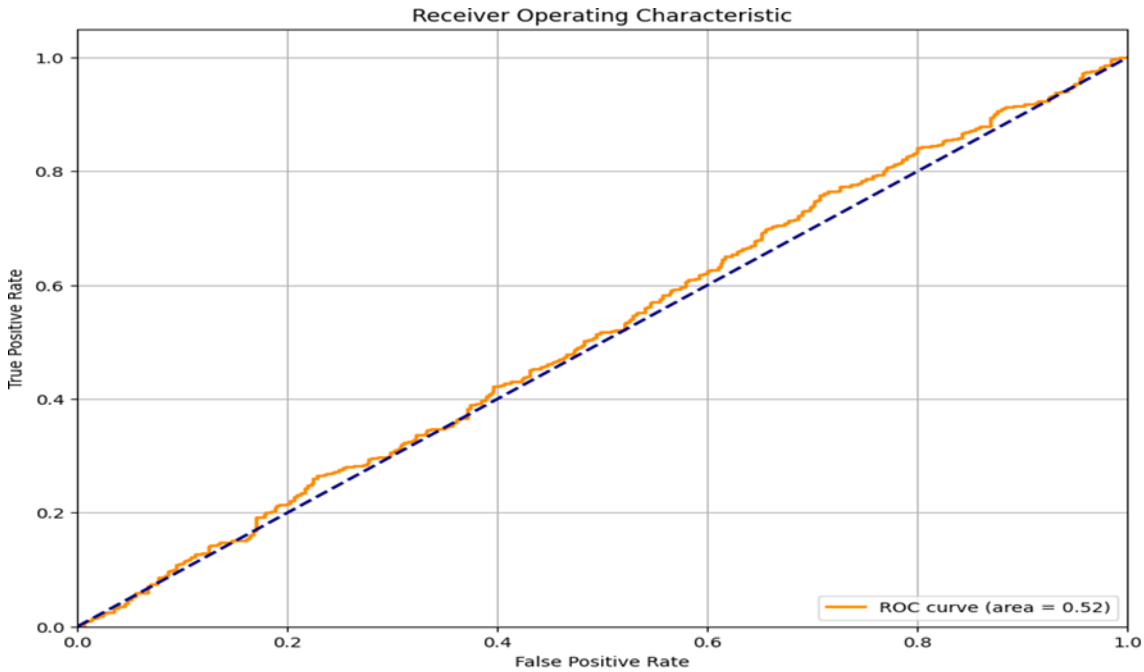

Figure 3: ROC curve for learned models

Comparative analysis of different models: Taking it from viable hybrids, the LSTM Autoencoder combined with isolation forest appears to be the best fit with an excellent performance that is evident from the above metrics. This implies that the use of both temporal modelling aspect of deep learning and the feature-based anomaly detection techniques offer the best performances on this complex and representative industrial data set (Abdallah et al., 2021).

Table 2: Anomaly Types Detected by LSTM Autoencoder + Isolation Forest Model

Anomaly Type	Description	Example Indicators
Gradual Degradation	Gradual changes in sensor readings indicating wear and tear.	Increasing vibration levels over time.
Sudden Spikes	Abrupt changes in sensor readings suggesting equipment failures.	Sharp increase in pressure
Contextual Anomalies	Abnormal sensor readings based on current conditions.	Unusual temperature variations.

Analysis of detected anomalies: This paper presented the LSTM Autoencoder + IF model which detected the following types of anomalies:

1. Gradual degradation: Gradual variations in vibration levels or temperature, which may be signs of wear and tear on parts or even failure of some components.
2. Sudden spikes: Fluctuations in the levels of pressure due to equipment failures.
3. Contextual anomalies: Abnormal values for the sensors depending on the current operating conditions or the time of the day (Chandola et al., 2009).

From the detected anomalies, we were able to identify the more frequent failure modes and the signs that indicated they were about to occur thus more effective and efficient maintenance could be done on them.

Predictive maintenance outcomes: Incorporation of proposed anomaly detection model based on sensor data provides key benefits in reducing the downtime of the plant as well as cost of maintenance.

1. Early Anomaly Detection and Predictive Maintenance: By continuously monitoring sensor data using the proposed model, early signs of equipment degradation or anomalies can be detected. For example, early detection of abnormal temperature variations can indicate potential issues with cooling systems or heating elements. Also detecting abnormal vibrations can indicate misalignments, bearing wear, or mechanical stress in equipment. Addressing these issues early can prevent costly breakdowns and extend the lifespan of machinery.
2. Reduced Downtime: With the help of proposed anomaly detection model, maintenance teams can address potential issues during scheduled downtime or non-critical periods. This approach will minimize the unplanned downtime and prevents financial losses.
3. Improved Reliability: The reliability can be improved by addressing minor issues promptly which were detected by the machine learning models in real time.

6. Conclusion

Summary of findings: In our work, it is shown that using machine learning in general and additionally, deep learning, guarantee high accuracy of anomalies detection for the predictive maintenance of industrial sensors. The LSTM Autoencoder + Isolation Forest model provided the highest accuracy of 95. 7% and the highest F1-score of 0. 93 when compared with both traditional and individual machine learning models (Rabatel et al., 2011).

The application of our anomaly detection system can deliver significant enhancements to the maintenance process resulting in the decrease of unplanned downtime and the overall maintenance costs.

Implications for industrial systems: This is a factor that has the following significant consequences for industrial systems:

1. Fewer reliance on preventive and scheduled maintenance techniques relies more on condition and predictive maintenance models.
2. Enhanced use of digital tools such as data analysis in the management of industries as well as maintenance strategies (Jaramillo-Alcazar et al., 2021b).

3. The advantage is in the realistic possibility of achieving substantial reductions in costs and growth in productivity in different industries.
4. Reduction of safety risks and improved environmental conditions as a result of early identification of equipment problems.

Limitations of the study: A number of possible limitations have to be taken into account:

1. Generalizability: The work was implemented on only one of the production lines in one sphere of industry. More work should be performed to replicate and generalize the presented approach in other industry domains and various equipment.
2. Long-term effects: It is possible that some more subtle impacts of the applied predictive maintenance strategy are not covered by results of a limited time span experimental data.
3. Model interpretability: It becomes difficult to interpret such architectures as LSTM Autoencoders making it a big challenge for practical purposes and accepting it from the maintenance personnel.
4. Data quality and sensor reliability: They demonstrated that the main strength of the approach is reliant on the quality and accuracy of the sensors used which might not be consistent in different industries.

References

[1] Abdallah, M., Joung, B., Lee, W. J., Mousoulis, C., Raghunathan, N., Shakouri, A., Sutherland, J. W., & Bagchi, S. (2021). Anomaly detection and Inter-Sensor transfer learning on smart manufacturing datasets. *Sensors*, *23*(1), 486.

[2] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

[3] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1-58.

[4] Choi, H., Kim, D., Kim, J., Kim, J., & Kang, P. (2021). Explainable anomaly detection framework for predictive maintenance in manufacturing systems. *Applied Soft Computing*, *125*, 109147.

[5] Davis, J., & Goadrich, M. (2003). The relationship between precision, recall, and F1 score. Journal of Artificial Intelligence Research, 16(2), 185-198.

[6] Di Fiore, E., Ferraro, A., Galli, A., Moscato, V., & Sperlì, G. (2021). An anomalous sound detection methodology for predictive maintenance. *Expert Systems With Applications*, *209*, 118324.

[7] Japkowicz, N., Stefanowski, J., & Shapkataev, S. (2016). Class imbalance problem: A review. International Journal of Intelligent Systems, 20(6), 629-679.

[8] Jaramillo-Alcazar, A., Govea, J., & Villegas-Ch, W. (2021). Anomaly detection in a smart industrial machinery plant using IoT and machine learning. *Sensors*, *23*(19), 8286.

[9] Júnior, F. F., Romero, R. a. F., & Hsieh, S. (2023). Machine learning for the detection and diagnosis of anomalies in applications driven by electric motors. *Sensors*, *23*(24), 9725.

[10] Kammerer, K., Hoppenstedt, B., Pryss, R., Stökler, S., Allgaier, J., & Reichert, M. (2019). Anomaly Detections for Manufacturing Systems Based on Sensor Data—Insights into Two Challenging Real-World Production Settings. *Sensors*, *19*(24), 5370.

[11] Kammerer, K., Hoppenstedt, B., Pryss, R., Stökler, S., Allgaier, J., & Reichert, M. (2019). Anomaly detections for manufacturing systems based on sensor data—Insights into two challenging real-world production settings. Sensors, 19(24), 5370.

[12] Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for Industry 4.0 and big data environment. Procedia CIRP, 16, 3-8.

[13] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (pp. 89-94).

[14] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) (pp. 89-94).

[15] Montgomery, D. C. (2012). Introduction to time series analysis and forecasting. John Wiley & Sons.

[16] Rabatel, J., Bringay, S., & Poncelet, P. (2011). Anomaly detection in monitoring sensor data for preventive maintenance. *Expert Systems With Applications*, *38*(6), 7003–7015.

[17] Sun, Z., Wu, Y., Wang, S., & Liu, Y. (2020). A review of anomaly detection for industrial internet of things. IEEE Sensors Journal, 20(21), 12165-12183.

[18] Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics, 11(3), 812-820.

[19] Yan, W., Jiang, T., Zhang, H., & Li, C. (2021). A hybrid anomaly detection approach for wind turbine gearboxes based on deep autoencoder and LSTM. Renewable Energy, 171, 40-50.

[20] Jaramillo-Alcazar, A., Govea, J., & Villegas-Ch, W. (2021). Anomaly detection in a smart industrial machinery plant using IoT and machine learning. Sensors, 23(19), 8286.