Journal of Electrical Engineering and Information Technologies, Vol. **10**, No. 1 pp. 19-24 (2025)
Article 228
Received: 01.05.2025
Accepted: 28.05.2025
In print: ISSN 2545–4250
On line: ISSN 2545–4269
UDC: 629.73.017:004.85.021

Original scientific paper

DETECTION OF ANOMALIES IN AIRCRAFT USING MACHINE LEARNING ALGORITHMS

Melanija Gerasimovska[1)]   Dushko Stavrov   Gorjan Nadzinski   Vesna Ojleska Latkoska

Ss. Cyril and Methodius University in Skopje, Faculty of Electrical Engineering and Information Technologies, 1000 Skopje, Republic of North Macedonia

[1)] melanija.gerasimovska378@gmail.com

Abstract: This paper investigates the application of machine learning (ML) algorithms for anomaly detection in aviation, focusing on predictive maintenance and improving safety through early fault identification. Time-series vibration sensor data from helicopters is used to evaluate four anomaly detection methods: Isolation Forest (IF), One-Class SVM, Local Outlier Factor (LOF), and Convolutional-Reconstruction Autoencoder (CRAE). Both supervised and unsupervised detection approaches are considered. One-Class SVM demonstrated the highest performance, achieving accuracy of 93.9% and F1-score of 93.6%, followed by LOF (91.9%) and Isolation Forest (86%). CRAE underperformed with F1-score of 66.5%, primarily due to minimal preprocessing. These results highlight the effectiveness of simpler ML models over complex deep learning architectures in environments with limited data and real-time constraints.

Keywords: anomaly detection, predictive maintenance, aircraft, machine learning, unsupervised learning

ДЕТЕКЦИЈА НА АНОМАЛИИ КАЈ ВОЗДУХОПЛОВИ СО ПРИМЕНА НА АЛГОРИТМИ ОД МАШИНСКО УЧЕЊЕ

Апстракт: Овој труд ја истражува примената на алгоритмите од машинско учење за откривање аномалии во воздухопловството, со фокус на предвидувачко одржување и подобрување на безбедноста преку рана идентификација на грешки. Во трудот се користат податоци од сензори за вибрации на хеликоптери за да се направи евалуација на четири методи за откривање на аномалии: изолирани шуми, машини со поддржувачки вектори за една класа, локален фактор за аномалии и конволуциско-реконструкционен автоенкодер. Разгледани се и пристапи на надгледувано и ненадгледувано откривање на аномалии. Машините со поддржувачки вектори за една класа покажаа највисоки перформанси, постигнувајќи точност од 93,9% и ф1-резултат од 93,6%, проследено со локален фактор за аномалии од 91,9% и изолирани шуми од 86%. Конволуциско-реконструкциониот автоенкодер покажа послаби резултати со ф1-резултат од 66,5% поради недостаток на претходна обемна обработка или модификации на податоците. Овие резултати ја истакнуваат ефикасноста на поедноставните модели на машинско учење во однос на сложените архитектури на длабоко учење во средини со ограничени податоци и ограничувања во реално време.

Клучни зборови: детекција на аномалии, предвидувачко одржување, воздухоплов, машинско учење, ненадгледувано учење.

I. INTRODUCTION

AIR transportation is becoming an increasingly common choice among people due to the possibility of reaching the desired destination in a much shorter time. However, as for any other type of transport, the safety of passengers comes first. Although the aviation's safety record is commendable, occasional rare incidents do occur, highlighting the need for continuous improvement of safety measures. Today, the aviation industry upholds rigorous safety standards that leave little room for concern. Still, there remains significant unexploited potential to further strengthen security, especially through application of contemporary technologies and methodologies like Artificial Intelligence (AI).

The process of maintenance of different sub-systems and parts contained in aircrafts is crucial in ensuring safety and longevity. The standard procedure of maintaining the aircraft in safe operational conditions is conducted by regular technical checks, while the airplane is on the ground. This is a predictive maintenance protocol carried out by technical personnel based on technical data sheets of the equipment, and feedback information from the electrical systems and pilots. The critical nature of the aircraft does not allow for reactive maintenance to be applied. With the use of Machine Learning (ML) algorithms, predictive maintenance can also be established based on data collected from real-time sensors, historical data of performance, and advanced analytics. Such predictive maintenance systems can be used to detect anomalies and predict possible failures [1] [2].

In addition to safety, the aviation industry faces a significant and ever-present challenge, i.e. the need to reduce costs and delays while maintaining and improving aircraft operations. Inadequate maintenance can lead to significant financial losses for airlines in the form of grounded aircrafts, passengers waiting, and flight cancellations.

With the increasing complexity of modern aircraft systems and the vast amount of data generated during flight operations, the need for robust anomaly detection mechanisms is substantial. Modern aircraft systems leverage advanced ML algorithms and data analysis techniques to detect anomalies, i.e. deviations from normal behavior, helping aviation professionals address potential issues before they escalate [3].

Anomalies may appear as irregular sensor readings, sudden deviations in flight parameters, or other unexpected behaviors. However, integrating ML methods into aircraft anomaly detection systems poses several challenges. These include real-time processing limitations, computational resource constraints, latency issues, and the need for seamless integration with existing aviation infrastructure.

Some papers analyzing the process of anomaly detection in aircrafts [4] rely on transformation of time series into images, which are then used as input features to Deep Learning (DL) algorithms. This approach led to quite promising results in the field of anomaly detection with accuracy of 92% and F1-score of 91%. On the other hand, paper [5] features an analysis of the same dataset, using LSTM (Long Short-Term Memory) and One-Class SVM (Support Vector Machines). The reported results indicate that One-Class SVM with simplest statistical features achieves better performance in respect to LSTM network. In [6], a high F1-score of 94% is reported when the Support Vector Data Description (SVDD) method is utilized to perform one-class classification by training exclusively on non-anomalous data. As a result, it defines a hypersphere in the feature space, where the boundary of the hypersphere serves as the threshold for detecting anomalies. In this approach, the features are extracted by Convolutional Neural Network (CNN) which, combined with SVDD method, completes an anomaly detection pipeline. In [7], Convolutional Autoencoder (CAE) is implemented for anomaly detection in vibration condition-based monitoring. Also, one can find results reported for anomaly detection by utilizing unsupervised anomaly detection methodology in flight data in [8]. To overcome the challenge of unavailable labeled data, they developed a Convolutional Variational Auto-Encoder (CVAE), which is an unsupervised deep generative model for anomaly detection in high-dimensional time-series data. By validating Yahoo's benchmark data and a case study involving identification of anomalies in commercial flight take-offs, it was demonstrated that CVAE surpasses both classic and deep learning-based approaches in terms of precision and recall for anomaly detection.

The main focus of this paper is a comparison of different ML algorithms for anomaly detection in aircrafts. The dataset [9] used is vibration measurements of helicopter's operating conditions. The idea is to establish a robust and stable model that will be capable of detecting anomalies in real time and distinguish between different types of anomalies with measurable margin of accuracy and confidentiality. A series of experiments with various algorithms, i.e. One-Class SVM (with and without SGD), Isolation Forest, Local Outlier Factor and Convolutional Reconstruction Autoencoder, are conducted to determine whether existing results in this field can be improved. Among ML methods tested, One-Class SVM achieved the best performance, with an F1-score of 93.6%, followed closely by Local Outlier Factor method at 91.9%. Both Isolation Forests and One-Class SVM with SGD (Stochastic Gradient Descent) yielded similar results, around 86%, while CRAE (Convolutional Reconstruction Autoencoder) method performed the worst, achieving a score of 66.5%.

This paper is structured as follows: Section 2 provides description and analysis of dataset used for training and testing ML algorithms; in Section 3, the algorithms, tests and results are discussed; and the conclusion is given in Section 4.

II. Data Description and Analysis

This paper analyzes data collected and published by Airbus SAS [9]. The data refer to vibration measurements of helicopters that have been tested during flight. The vibration data are obtained with multiple accelerometers placed at different locations on the helicopter, in different directions (longitudinal, vertical, lateral), to measure vibration levels under different operating conditions of the helicopter. It is important to note that we are dealing with a sensitive problem, as data are limited and rare.

The dataset consists of 2271 one-dimensional time-series, each lasting 1 minute. The time-series are recorded during different flights and are sampled with frequency of 1024 Hz. Moreover, the dataset is already split into training and test subsets, noting here that there are no anomalous time-series in the training subset. The test subset contains both anomalous and non-anomalous time-series with anomalous time-series being labeled without pinpointing the exact moment(s) in time when an anomaly occurs. The training and test data are composed of 1677 and 594 1-minute sequences, respectively.

Each 1-minute sequence has 61440 data points. All anomaly-free sequences encompassed in the training set were used to analyze the normal behavior of accelerometer data and to train ML algorithms to what is considered normal anomaly-free behavior. Test data are evenly split, with 50% (297) representing anomalous time-series and the other 50% representing non-anomalous time-series. In total, the entire dataset contains 13% anomalous and 87% non-anomalous time sequences. An illustration of non-anomalous and anomalous data sequence is given in Figure 1 and Figure 2, respectively.

The main idea is to design a ML model based on non-anomalous training data which is then capable to differentiate between input time-sequences that contain anomalies and the ones that are anomaly-free. The dataset analysis showed there are no missing values in time sequences. Value ranges of time-series for training data, non-anomalous test data and anomalous test data are given in Table 1. There is a noticeable difference in ranges between data with and without anomalies, which is here hypothesized to be a direct consequence of the presence of anomalies in the time-series. Thus, it is expected that the significant gap between min and max values of anomaly and anomaly free time-series will be exploited by ML algorithms to distinguish between the two classes of data: anomalous and non-anomalous.

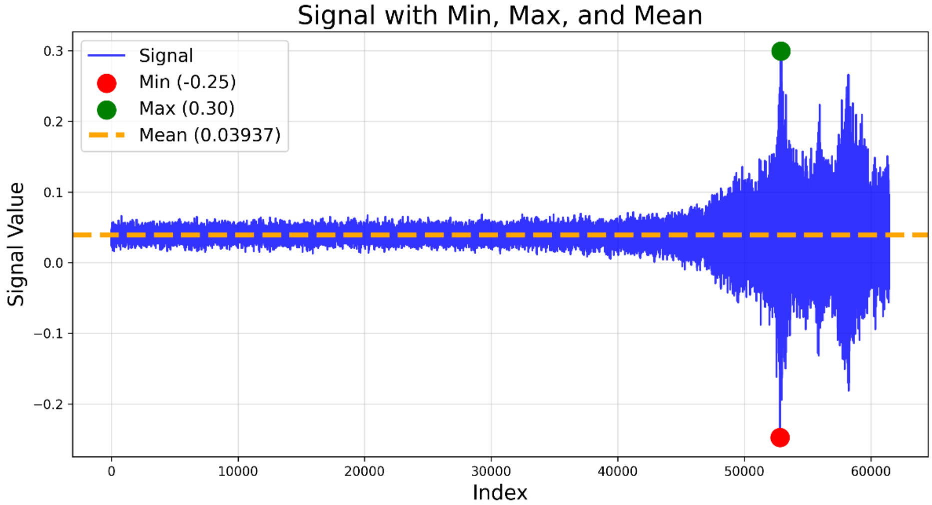

Fig. 1. Non-anomalous sequence consisted of data points collected over 1 minute with min and max values of -0.25 and 0.30, respectively

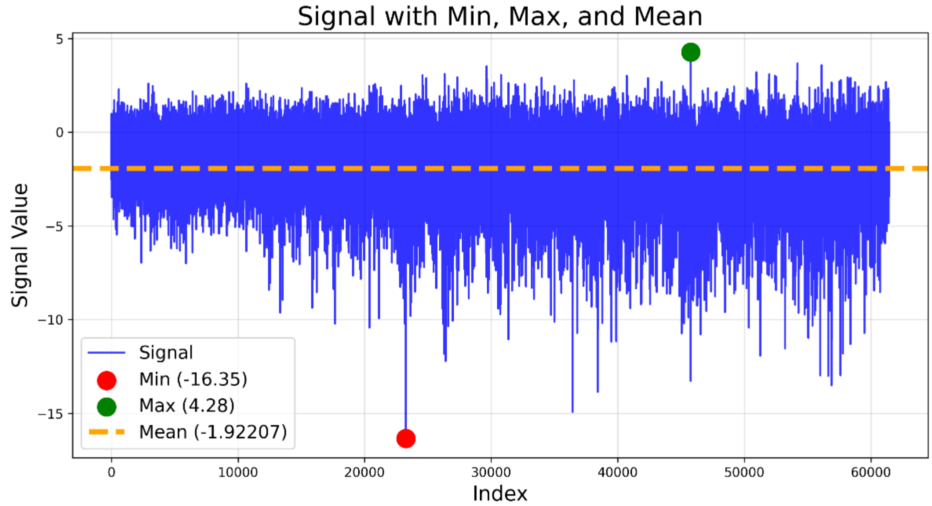

Fig. 2. Anomalous sequence consisted of data points collected over 1 minute, with min and max values of -16.35 and 4.28, respectively

An analysis of the variance of the time-sequences was also performed on the entire dataset, resulting in observance of time-series that are anomaly-free, but of very high variance corresponding to anomalous signals.

TABLE I

MIN AND MAX VALUES OF TRAINING AND TEST DATA

Dataset	Min value	Max value
Training data	-19.02	17.91
Non-anomalous test data	-4.89	4.61
Anomalous test data	-35.7	11.2

This was addressed by applying mean filtering with various window lengths. The primary purpose of mean filtering is to reduce random noise and smooth out fluctuations in the data, making it easier to identify underlying patterns. By averaging each data point with its neighboring values, mean filtering helps to suppress short-term irregularities that may be caused by noise, while preserving more significant trends and abrupt changes that are likely to be true anomalies. Finally, a window length of 10 was employed providing proper decrease in the variance, while avoiding excessive smoothing which could potentially mask true anomalies. As a result of filtering, the variance of high-variance signals was decreased, ensuring retention of the data containing critical information about the anomalies.

A. Feature Engineering

An important consideration when designing features for anomaly detection is the choice of algorithm used for training and testing. Certain algorithms, such as deep learning-based unsupervised approaches (e.g., Autoencoders, Generative Adversarial Networks) [10] [11], directly process raw data without requiring explicit feature engineering, as they have innate capability to learn autonomously. Others, such as traditional ML methods (e.g., Support Vector Machines, Random Forests, or clustering algorithms like K-means) [12], rely heavily on well-crafted features to extract meaningful insights from the dataset which, in turn, enable the algorithm to better identify patterns and anomalies. For the purpose of the analysis carried out in this paper, several algorithms were employed for anomaly detection: Isolation Forest, One-Class SVM, One-Class SVM with SGD, Local Outlier Factor, and Convolutional-Reconstruction Autoencoder. Features were not engineered for Isolation Forest and Convolutional-Reconstruction Autoencoder, as these algorithms work well directly with raw data. For algorithms such as One-Class SVM and Local Outlier Factor, statistical features were designed to enhance performance. These features include the mean, standard deviation, kurtosis, skewness, quantiles, and median, extracted over varying window lengths such as 1024, 512, 341, 256, 103, 51, and 26 data points. Each window length corresponds to specific time duration, with 1024 data points representing 1000 milliseconds (1 second), and the others scaling accordingly. When testing different window lengths, the objective is to identify the optimal window size that yields the best performance for the specific anomaly detection task. Since a sampling frequency of 1024 Hz results in 1024 data points per second, it is crucial to extract features from data windows that span 1 second or less. This approach ensures timely detection of anomalies, as shorter windows enable faster response times while still capturing meaningful statistical patterns within the data. Additionally, some experiments were conducted using only variance as feature, while others combined statistical features with those derived from Fourier transform and Principal Component Analysis (PCA). Fourier transform was particularly useful for detecting the frequencies of anomalous signals. Namely, by analyzing the frequency spectrum of the anomalous time-series a pattern was detected where some of the anomalous time-series contain a dominant frequency component between 50 and 60 Hz. On the other hand, PCA helps in isolating the dominant patterns or anomalous frequency behaviors within the signal. Since anomalous signals often exhibit distinct frequency characteristics, PCA can highlight these variations by emphasizing the components where anomalies cause significant deviations. Fourier transform of non-anomalous and anomalous sequence is illustrated on Figure 3 and Figure 4, respectively. As it can be observed, the anomalous sequence contains frequency components with notable magnitudes, while the magnitudes of frequency components in the anomaly-free sequence are contained within a considerably narrower, negligible range.

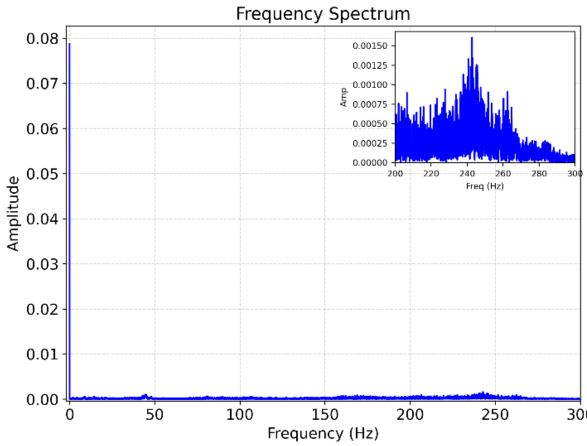

Fig. 3. Frequency domain of a time-sequence without anomalies

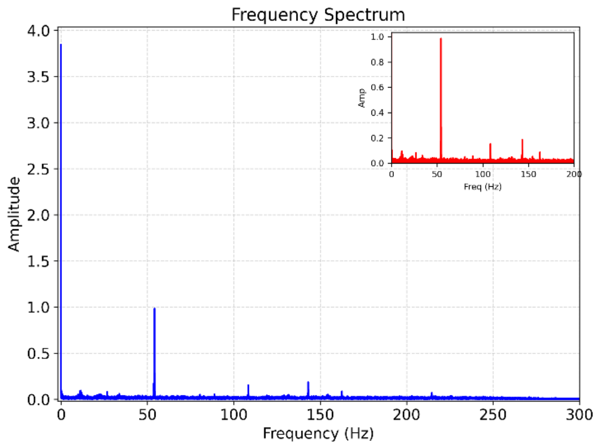

Fig. 4. Frequency domain of a time-sequence with anomalies

III. TESTS AND RESULTS

A comprehensive comparison is given in Table 2 showing the performance of five ML algorithms on the test set. Their performance is evaluated by using the following metrics: Accuracy, Precision, Recall and F1-score, allowing for an in-depth analysis of their strengths and weaknesses. The equations used for calculating these metrics are given in equations (1), (2), (3) and (4), respectively. The metric values are determined by combining True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) instances, and total number of samples (N). It is important to note that tests were made using mean filtering with a window length of 10 instances (data points) and without mean filtering, in order to find out which one has greater impact on the performance of the algorithms. The same tests are repeated for both filtered and unfiltered datasets, with same statistical features extracted, the same parameter values tried, etc.

$$Accuracy = \frac{TP + TN}{N} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

Among the algorithms tested, One-Class SVM [13] performed better than the others, achieving the highest accuracy (93.9%) and F1-score (93.6%), with distinguishable precision (97%) and the third best recall (89.8%). These results were achieved by setting the parameters nu: 0.000003, gamma: scale, kernel: rbf, degree: 1, and window length: 512, and without mean filtering the dataset. All this show that One-Class SVM provides a strong tradeoff between detecting anomalies and minimizing FP and FN, making it the most reliable choice for anomaly detection in the helicopter example.

TABLE II
METRIC VALUES OF THE ALGORITHMS ON TEST SET

Algorithm	Accuracy	Precision	Recall	F1-score
Isolation Forest	87%	86%	87%	86%
One-Class SVM	93.9%	97%	89.8%	93.6%
One-Class SVM with SGD	87.7%	97.8%	77%	86.2%
Local Outlier Factor	92%	93.7%	90.2%	91.9%
CRAE	74.9%	100%	49.8%	66.5%

Figure 5 illustrates the confusion matrix of the One-Class SVM model.

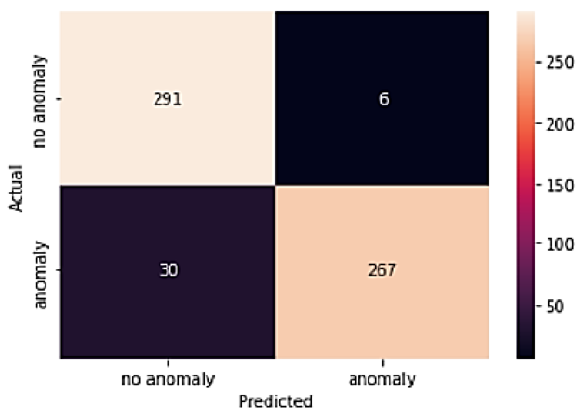

Fig. 5. Confusion matrix of One-Class SVM

According to the confusion matrix, the model correctly identified TP=291 anomaly-free sequences. It incorrectly classified FP=6 normal sequences as anomalies (false alarm). The model missed FN=30 anomaly sequences, incorrectly classifying them as normal, and it correctly detected TN=267 anomaly sequences. This model has achieved commendable performance overall; however, the high number of FN instances should be carefully considered when deciding on real-world implementation.

The Local Outlier Factor (LOF) [14] algorithm closely follows the One-Class SVM performance, with an accuracy of 92%, precision of 93.7%, recall of 90.2%, and F1-score of 91.9%. The best results are obtained for window length 256, neighbors: 20, contamination: 0.0003, and mean filtering applied. While slightly behind the One-Class SVM, LOF offers comparable performance and serves as robust alternative, primarily due to its greater interpretability and computational efficiency.

The Isolation Forest [15] achieves the third best result overall, with balanced values across all metrics (accuracy: 87%, precision: 86%, recall: 87%, and F1-score: 86%). This result was achieved with setting contamination parameter to 0.02 and mean filtering applied. It can be considered a reliable option for simpler, non-critical applications and tasks requiring minimal tuning. Similarly, the One-Class SVM with SGD, when mean filtering is applied, prioritizes precision (97.8%) but sacrifices recall (77%), resulting in an overall F1-score of 86.2%. This algorithm is better suited for scenarios where avoiding FP is critical. However, in this case, minimizing FN is more important, as accurate anomaly detection is paramount.

In contrast, CRAE (Convolutional Reconstruction Autoencoder) [16] delivers the weakest overall performance, with an accuracy of 74.9%, recall of only 49.8%, and the lowest F1-score of 66.5% for window length 512 and mean filtering applied. While it achieves precision of 100%, it is still not suitable for use in our case. This could be due to its being a complex approach that requires additional adjustments to make it more suitable for anomaly detection in aircraft.

In summary, the results shown in Table 2 suggest that One-Class SVM is potentially the most suitable choice for integration into a system framework for accurately detecting critical anomalies, such as those occurring in airplanes. Its strong balance across performance metrics and proven reliability provide a suitable option for safety-critical applications. LOF algorithm is a compelling alternative due to its robust performance, ease of implementation on hardware, and interpretability, which are particularly advantageous in resource-constrained environments. Ultimately, the choice of algorithm should be guided by the specific requirements of the application, including priorities such as precision, recall, computational efficiency, and system constraints

IV. CONCLUSION

This paper focuses on anomaly detection in helicopters using ML algorithms, tackling challenges such as limited data availability and the need for real-time implementation on aviation hardware. By leveraging ML techniques that are less computationally intensive than deep learning, this study efficiently balances resource constraints with performance. Contrary to common assumptions, simpler ML models (like One-Class SVM and LOF) outperformed more complex DL algorithms (like CRAE), demonstrating that algorithm complexity does not always correlate with efficacy. The results highlight that simpler, well-tuned ML models can achieve comparable results with less computational overhead, making them suitable for aviation systems.

This study emphasizes the importance of minimizing false negatives in safety-critical systems like helicopters, as missed anomalies can potentially lead to catastrophic outcomes. While no algorithm guarantees perfect detection, One-Class SVM achieved the best balance between reliability, precision, and recall, making it the most suitable choice for aviation. The findings also provide potential directions for future research, such as enhancing preprocessing for DL models, exploring neural network alternatives, and improving anomaly localization techniques.

To summarize, the findings of this study reinforce the value of adopting efficient ML-based anomaly detection frameworks in aviation to support predictive maintenance and improve flight safety.

REFERENCES

[1] F.A. Adryan, K.W. Sastra. "Predictive Maintenance for Aircraft Engine Using Machine Learning: Trends and Challenges", 30.06.2021.

[2] Hermawan, A. P., Kim, D-S., Lee, M-J. "Predictive Maintenance of Aircraft Engine Using Deep Learning Technique". International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 1296-1298.

[3] Charu C. Aggarwal. "Outlier Analysis," IBM T. J. Watson Research Center Yorktown Heights, second edition, New York, 25.11.2016.

[4] Gabriel Rodriguez Garcia, Gabriel Michau, Mélanie Ducoffe, Jayant Sen Gupta, Olga Fink. "Temporal Signals to Images: Monitoring the Condition of Industrial Assets with Deep Learning Image Processing Algorithms", 26.02.2021.

[5] Kilian Vos, Zhongxiao Peng, Christopher Jenkins, Md Rifat Shahriar, Pietro Borghesani, Wenyi Wang. "Vibration-Based Anomaly Detection using LSTM/SVM Approaches", 15.04.2022.

[6] Chenyu Liu, Konstantinos Gryllias. "A Deep Support Vector Data Description Method for Anomaly Detection in Helicopters", 29.06.2021.

[7] Vaibhav Malviya, Indrani Mukherjee, Siddharth Tallur. "Edge-Compatible Convolutional Autoencoder Implemented on FPGA for Anomaly Detection in Vibration Condition-Based Monitoring", 16.04.2022.

[8] Milad Memarzadeh, Bryan Matthews, Ilya Avrekh. "Unsupervised Anomaly Detection in Flight Data Using Convolutional Variational Auto-Encoder", 08.08.2020.

[9] ETH Zurich, AIRBUS SAS, Airbus Helicopter Accelerometer Dataset, 2022. Available at: https://www.researchcollection.ethz.ch/handle/20.500.11850/415151.

[10] Mikael Sabuhi, Ming (Chloe) Zhou, Cor-Paul Bezemer, Petr Musilek. "Applications of Generative Adversarial Networks in Anomaly Detection: A Systematic Literature Review", 26.10.2021

[11] Kamal Berahmand, Fatemeh Daneshfar, Elaheh Sadat Salehi, Yuefeng Li, Yue Xu. "Autoencoders and Their Applications in Machine Learning: A Survey", 03.02.2024.

[12] Iqbal H. Sarker. "Machine Learning: Algorithms, Real-World Applications and Research Directions", 22.03.2021

[13] Bernhard Schölkopf, Robert C. Williamson, Alex J. Smola, John Shawe-Taylor, John C. Platt. "Support Vector Method for Novelty Detection", January 1999.

[14] Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. "LOF: Identifying Density-Based Local Outliers", 2000.

[15] Fei Tony Liu, Kai Ming Ting, Zhi-Hua Zhou. "Isolation Forest", January 2009.

[16] Masci, J., Meier, U., Ciresan, D., & Schmidhuber, J. "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction", 2011.