Journal of Engineering and Applied Research (JEAR)

Technical and Vocational University

Winter & Spring 2024, Vol. 1, No. 1, p. 183-196

Journal Homepage: https://jear.tvu.ac.ir/

doi:10.48301/JEAR.2024.447671.1022

TVU

Reducing Air Pressure System Repair Costs in Scania Trucks through Deep Learning

Kazem Taghandiki[1], Morteza Dallakehnejad[2*], Hossein Rahimi Asiabaraki[3]

[1]Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran.
[2,3]Department of Mechanical Engineering, Technical and Vocational University (TVU), Tehran, Iran.

ARTICLE INFO

Article Type:
Original Research

Received: 03.09.2024
Revised: 04.09.2024
Accepted: 05.05.2024

Keyword:
Deep Learning
Air Pressure System
Preprocessing
Feature Selection
Dataset

***Corresponding Author:**
Morteza Dallakehnejad
Email:
mdallakehnejad@tvu.ac.ir

ABSTRACT

Air pressure systems play a fundamental role in Scania trucks because the proper functioning of the brake and gear shifting systems of these vehicles relies on the health of the air pressure system. The presence of sensors in the air pressure system gathers various information about its status, which can be stored and analyzed in the form of datasets. Using machine learning algorithms to detect faults in the air pressure system prevents manual inspection at different time intervals, thus preventing material and time costs. Many efforts have been made to detect faults in the air pressure system through the collected datasets from sensors of the various components of Scania trucks using traditional machine learning algorithms such as decision trees, KNN, Random Forest and SVM, but they still lack sufficient accuracy and speed in data processing. However, since the number of records collected at different intervals by the Electronic Control Unit (ECU) is very high, novel machine learning algorithms like deep learning can be used to increase accuracy and speed in detecting faults in Scania truck air pressure systems. In this article, feature selection and deep learning algorithms have been used to detect and predict faults in the air pressure system of heavy trucks. The results and observations showed that the output of the evaluation parameters of the deep learning algorithm has an accuracy rate of 98.66%, a recall rate of 63.47%, and a f-measure rate of 68.99%.

E-ISSN: 3060-6357

Introduction

Scania trucks made in Sweden are usually expensive and have very heavy engines. These trucks are mostly sold in countries such as America, United Kingdom, Canada, Russia, and Singapore. Instead of using hydraulic fluid, Scania trucks use an air pressure system for proper operation of the brake and gear shifting system [1]. Since Scania trucks are used in many industries and in long journeys, monitoring the air pressure system of Scania trucks is vital for the proper functioning of the braking and gear shifting system [2].

Proper monitoring of the air pressure system prevents the failure of other parts and increases the cost of Scania trucks. In recent years, the automotive industry has sought to reduce operating costs and maintenance, and increase customer satisfaction by using digital tools and software. For instance, the presence of ECU hardware parts in Scania trucks collects information from different parts under the name of a data set. By analyzing and examining this data set, researchers can provide solutions to quickly diagnose air pressure system failure in heavy trucks [3; 4].

Many research studies [5-12] have been conducted to identify and quickly diagnose the failure of the air pressure system in Scania trucks. However, these studies are mostly focused on using traditional machine learning algorithms such as decision trees, logistic regression, SVM, KNN, and Random Forest [13-15]. As the volume of data collected from ECU hardware components increases, these algorithms may no longer provide the desired accuracy [16].

The APS dataset (air pressure system failures in Scania trucks) is used in the proposed approach. It collects data from ECUs in Scania trucks and has two training and testing sections with a record number of 60,000 and 16,000 [8], respectively. In the proposed approach, first, by using the three feature selection algorithms of Correlation, Information Gain and SVM, the most important and weighted features were identified among the 170 features in the APS dataset, and 21 features were identified as important features. Since many of the selected features had noise values, in the next step, using the Replace All Missing operator, the process of repairing the noise feature was carried out automatically. Then, using the Deep Learning algorithm in addition to the activation function, the process of training the deep learning model was carried out through the pre-processed dataset [17].

Finally, by evaluating the deep learning model trained through the experimental data set, it was found that the use of the deep learning algorithm has an accuracy of 98.66%. In the second part of the article, the steps taken to diagnose the failure of the air pressure system of Scania trucks more quickly is examined. In section 3, the implementation process of the proposed approach is described. Then, in section 4, the observations and results of the proposed approach are examined. Finally, the conclusion of the proposed approach is discussed in section 5.

Related work

There has been an increasing interest in using deep learning algorithms for optimizing the repair cost of air pressure systems in heavy-duty trucks. In one study, researchers used a neural network to predict the remaining useful life of air compressors in heavy-duty trucks, resulting in a 60% reduction in maintenance costs [18; 19]. Another study proposed a deep reinforcement learning algorithm to optimize the scheduling of maintenance tasks for truck fleets, which included air pressure system maintenance. Additionally, other research has focused on using machine learning techniques to detect and diagnose faults in air pressure systems, which could lead to more efficient repairs and cost savings. Overall, these studies demonstrate the potential of deep learning algorithms for minimizing the repair cost of air pressure systems in Scania trucks.

Selvi et al. [1] used the APS data set to detect and prevent the failure of the air pressure system of Scania heavy trucks; they first used the sk-learn library and sampling algorithms such as under-sampling, over- sampling and SMOTE to balance the APS dataset. Then, using traditional machine learning algorithms such as SVM, Logistic Regression, Random Forest, KNN, SGD, Decision Tree and Naive Bayes, they identified and predicted failure in the APS balanced data set. Their approach had the highest accuracy in Random Forest and KNN algorithms of 98%.

Cerqueira et al. [2] used the APS dataset to detect and prevent the failure of the air pressure system of Scania heavy trucks. First, they solved the two basic problems of lack of balance and the existence of noise characteristics in their selected dataset. They used the SMOTE algorithm to balance the number of dataset records to solve the problem of the APS dataset not being balanced, and also used 3 noise feature analysis algorithms to remove features with noisy values. Finally, they used the XGBoost and Random Forest algorithm to investigate the cost reduction in the air pressure system of heavy trucks where the XGBoost algorithm showed the best performance in the output results.

Rawat [3] used machine learning methods to predict the failure of parts related to the air pressure system in Scania trucks (APS dataset). He specifically focused on the pre-processing processes of his selected dataset where he first identified and removed the noisy features of the initial data set and then balanced his dataset using different balancing algorithms such as SMOTE. Finally, by applying the normalization and standardization processes, he normalized the APS dataset feature values. After applying pre-processing processes using Random Forest, Naive Bayes, SVM, KNN, and logistic regression algorithms, he predicted the failure of parts related to the air pressure system in heavy trucks, the results of which demonstrated an accuracy of 99% for the Random Forest algorithm.

Costa et al. [6] sought to provide an approach to identify and diagnose air pressure system failures using machine learning algorithms on the APS dataset. After applying pre-processing processes using Random Forest, decision tree, SVM, KNN and logistic regression algorithms, they predicted the failure of parts related to the air pressure system in heavy trucks; their results demonstrated the highest accuracy (92.5%) for the Random Forest algorithm.

Rafsunjani et al. [11] have sought to improve the results of machine learning algorithms by using pre-processing methods on the APS dataset. First, they solved the problem of noise features in the APS dataset by using 5 algorithms: Expectation Maximization, Mean Imputation, Soft Impute, MICE, and Iterative SVD. They have also used the under-sampling algorithm to solve the challenge of the imbalance of the data set. Finally, the pre-processed data set was analyzed using the 5 machine learning algorithms of Naive Bayes, SVM, KNN, Random Forest and Gradient Boosted Tree to identify and diagnose the failure of the air pressure system. Their results showed the highest accuracy (98%) for the Random Forest algorithm.

The proposed approach

The proposed system consists of 4 steps; Figure 1 shows the implementation process of the proposed approach. In the following sections each of the steps in Figure 1 are explained in detail.

Select dataset

In this research, the APS dataset available on the UCI website was used [20]. This dataset was collected by Mr. Lingreen and Bitos in 2016 and provided to researchers. The APS dataset is derived from data collected by sensors of the air pressure system of heavy trucks (Scania), which play an important role in determining the optimal performance of the braking and gear shifting system. The APS dataset has two training and testing sections; the training section of this dataset is used to train the final model and the testing section is used to evaluate the final model obtained.

1- The APS training segment dataset has 170 input features and 60,000 records. 59000 records have negative labels and 1000 records have positive labels.
2- APS experimental part dataset has 170 input features and 16,000 records. 15625 records have negative labels and 375 records have positive labels.

The 170 specified features do not include the number of input features and the tag feature (class). Records with a positive label (pos) indicate failure in the air pressure system in heavy trucks and records with a negative label (neg) indicate no failure of the air pressure system in heavy trucks.

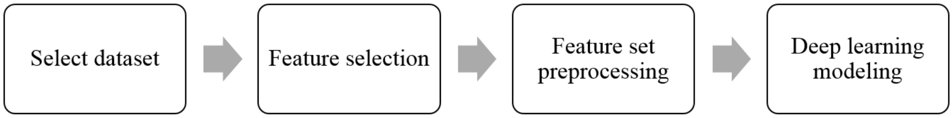

Figure 1. Process of the proposed approach.

Feature selection

In this step, the selection and weighting of important features from the APS dataset were carried out separately with 3 feature selection algorithms, information gain, correlation and SVM. A separate list was prepared for each of the feature selection algorithms where the the important and common features in the output of each of these 3 algorithms (3 lists produced) as important features and those with more weight were selected. Therefore, among the 170 features of the APS educational dataset, only 21 features were identified as important features, which will be selected as input features in the pre-processing stage. The process of this step is shown in Figure 2 using Rapid Miner software [21].

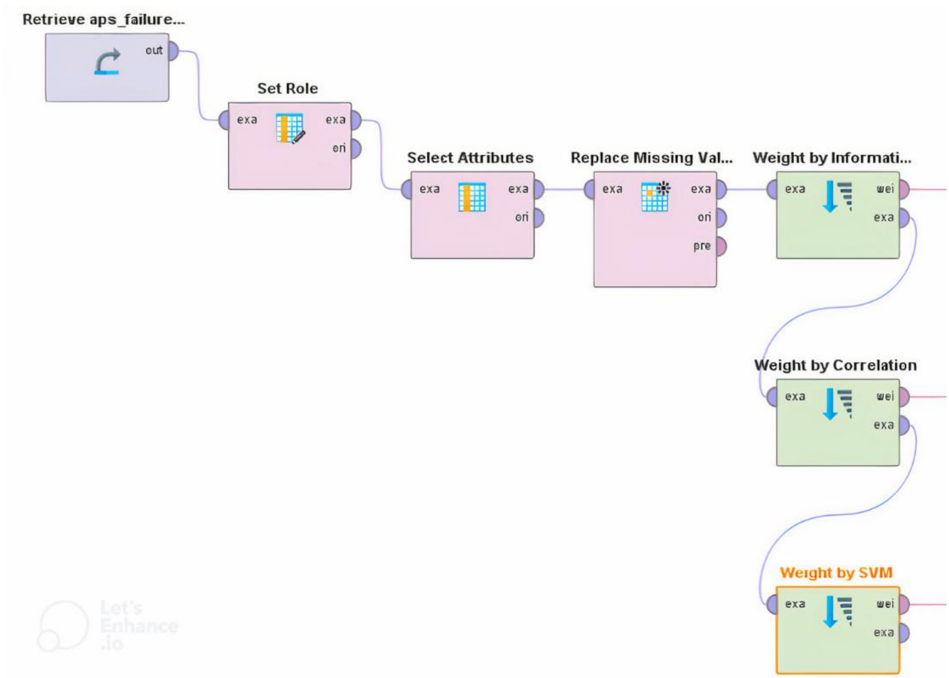

Figure 2. Selection process.

Feature set preprocessing

Since some of the 21 features selected in the previous step (feature selection) have noise values and the label feature values (class) are incompatible with the selection approach, it is necessary to take measures to resolve these types of inconsistencies, which is called pre-processing. Preprocessing is used in machine learning to improve the accuracy of a final model and improve the quality and quantity of the data. Therefore, at this stage, 4 pre-processing processes mentioned below were carried out [22].

Convert label attribute to binomial data type

The APS dataset tag feature is named as class in both the training and testing datasets. Considering that the values of the class attribute are only two values, neg and pos, the data type of this attribute is changed from polynomial to binomial by using the Nominal to Binominal operator.

Specify label attribute

When adding the training and testing data set, the label (class) feature must be recognized and therefore, the Set Role operator is used to specify the class feature as the label (class) feature.

Selection of input features

According to the identification of 21 important features in the feature selection stage (previous stage) using the Selected Attribute operator, these 21 features are selected as input features and the rest of the features are removed from the model building process. Table 1 shows the list of 21 selected features.

Identification and quantification of noise features

Because most of the features of the APS dataset have noise values, it was determined that all of them have limited noise values by examining the 21 input features. Therefore, by using the Replace All Missing operator, the average values in each of the numerical features of the dataset were automatically replaced by the noise values of that feature. Table 2 shows the number of noise values in each of the 21 selected features.

Table 1. Selected features.

Attribute	ci_000	ay_008	ba_005	ee_005	cc_000	cn_004	ck_000
Attribute	aq_000	cs_004	by_000	ba_004	bj_000	bb_000	bu_000
Attribute	bv_000	cq_000	bg_000	ah_000	bx_000	an_000	ao_000

Table 2. Number of noise values of selected features.

Attribute	ah_000	an_000	ao_000	aq_000	ay_008	ba_004	ba_005
Missing	645	642	589	589	671	688	688
Attribute	bb_000	bg_000	bj_000	bu_000	bv_000	bx_000	by_000
Missing	645	642	589	691	691	3257	473
Attribute	cc_000	ci_000	ck_000	cn_004	cq_000	cs_004	ee_005
Missing	3255	338	338	687	691	669	671

Figure 3 shows the set of pre-processing processes on two parts of the training and testing datasets.

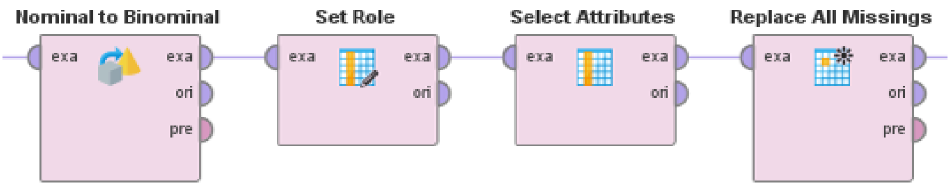

Figure 3. The set of pre-processing operations on the training and test datasets.

Deep learning modeling

In the third phase of the proposed approach, a deep learning model was built from the pre-processed training data set using the deep learning operator and then the activation function. The deep learning model has 4 layers (one input layer, two hidden layers, and one output layer) where the size of the input layer is 21, the size of the two hidden layers is 50, and the size of the output layer is 2. Figure 4 shows the weighting of the 21 input features selected by the deep learning algorithm.

The next step was to check whether the trained model has the desired and ideal accuracy or not. Therefore, using the Apply Model operator, the output of the trained deep learning algorithm model was received as an input, and the process of prediction and evaluation of the obtained deep learning model was performed on the test dataset (with 16,000 records).

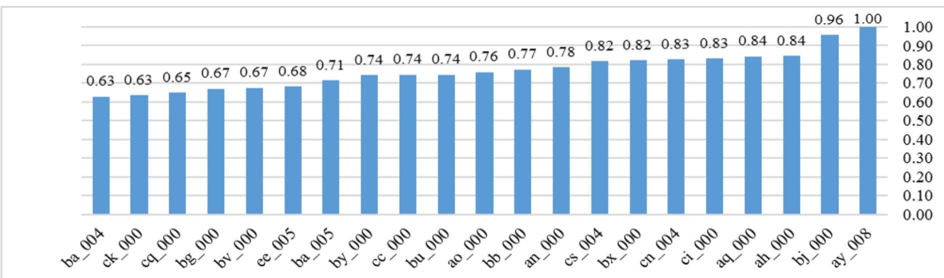

Figure 4. Weights assigned to input features in deep learning model

Figure 5 shows how the trained deep learning model performed the label assignment and prediction process for the records of the test dataset. The predication column shows the label assigned by the deep learning model and the class column shows the actual label of the record [23].

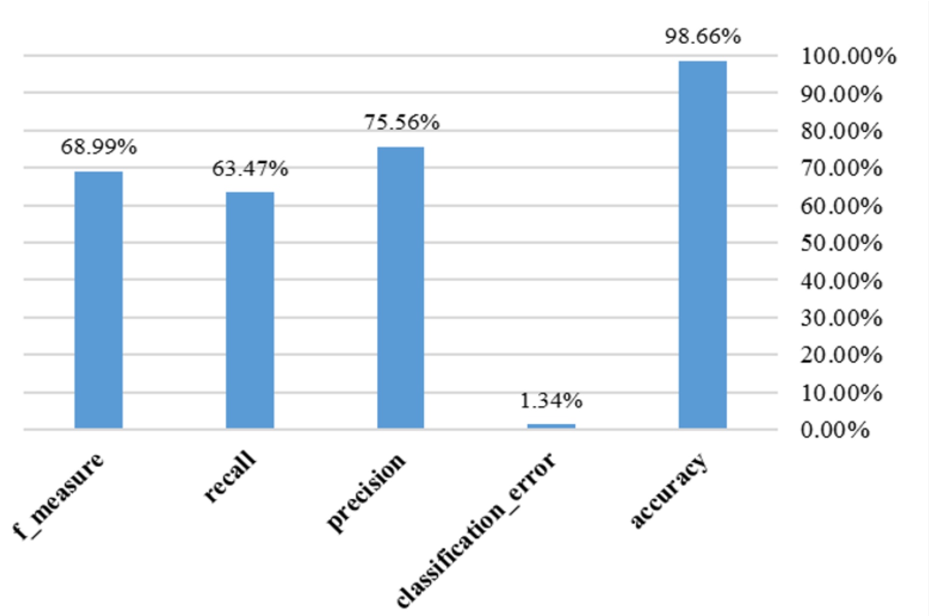

Figure 5. Deep learning model evaluation.

Figure 6 shows the complete process of implementing the deep learning model in the Rapid Miner machine learning software. Finally, using the Performance operator, the results obtained by the machine learning model for the records of the experimental data set were evaluated, which is fully explained in Section 4.

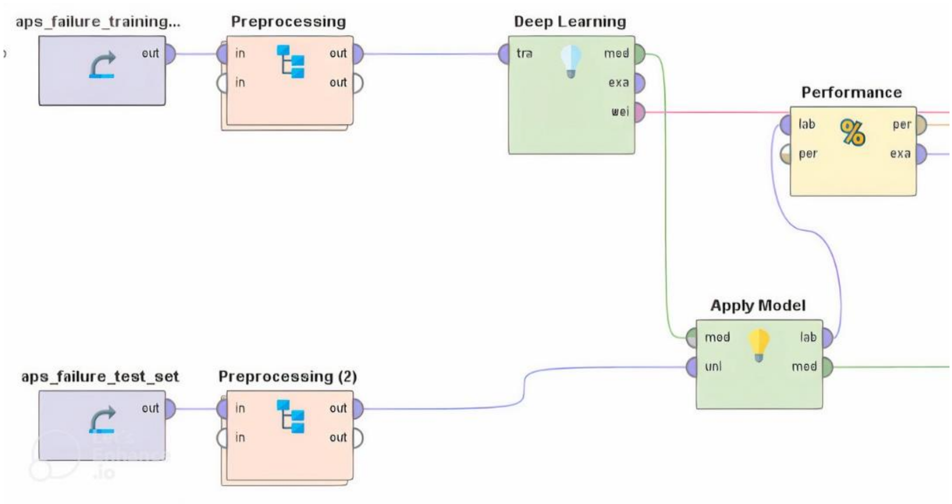

Figure 6. The complete process of implementing deep learning model in Rapid Miner machine learning software.

Observation and results

Based on the labels predicted by the trained deep learning model and the actual labels of each record of the test dataset, the clutter matrix [24] can be created, after which it is possible to obtain the evaluation parameters of deep learning models. Figure 7 shows the Confusion matrix obtained from the results of the deep learning model on the experimental dataset.

		Prediction classes			TP	238
		Neg	Pos		TN	15548
Real classes	Neg	15548	77	15625	FP	77
	Pos	137	238	375	FN	137
		15685	315			

Figure 7. Confusion matrix obtained from deep learning model results.

The final model of deep learning can be evaluated by specifying the parameters of the Confusion Matrix. Figure 8 presents the results of 5 evaluation criteria of the final deep learning model.

class	prediction(cl…
neg	neg
neg	neg
neg	neg
neg	neg
neg	neg
pos	pos

Figure 8. Label prediction of test dataset records by trained deep learning model.

In addition, the output of the obtained model was compared with the results of 5 traditional machine learning algorithms (SVN, Naive Bayes, Decision Tree, Random Forest, KNN), the observations obtained are shown in Table 3.

Table 3. Evaluation parameters of traditional machine learning algorithms.

ML Algorithms	Accuracy	Error rate	Precision	Recall	F Measure	Run Time (s)
SVM	98.18	1.82	81.34	29.07	42.83	300
Naive Bayse	96.90	3.10	42.11	86.13	56.57	25
Decision Tree	97.95	2.05	83.10	15.73	26.46	22
Random Forest	98	2	86.67	17.33	28.89	100
KNN	98.33	1.67	77.31	44.53	56.51	250
Deep Learning	98.66	1.34	75.56	63.47	68.99	41

The main advantage of this research compared to previous studies is the practical implementation of preprocessing operations in the deep learning model construction process. In the present research, two preprocessing operations were performed: one aimed at identifying and dealing with noisy features, and the other involved in converting some features from nominal data type to binominal. As a result, there was a relative improvement in execution speed, improvement in TP, TN, FP, and FN parameters as well as an increase in accuracy, a decrease in error rate, and an increase in the average harmonic mean compared to previous studies.

Conclusion

Air pressure systems play a crucial role in Scania trucks as the proper functioning of Scania vehicles' brake and gear shifting system depends on the health of the air pressure system. The presence of sensors in the air pressure system allows collecting various information on its status, which can be stored and analyzed in various datasets. Using machine learning algorithms to detect faults in the air pressure system prevents manual inspections at different time intervals, thus avoiding excessive material and time costs. In the current study, deep learning algorithms and feature selection were used to detect and predict faults in the air pressure system of heavy trucks using the APSFST dataset. The workflow involved identifying noisy features, applying preprocessing to them, converting some nominal features into binominal type to ensure the final model's acceptable accuracy. Finally, a model for detecting and predicting faults in the air pressure system of heavy trucks was developed using deep learning algorithms. The results and observations demonstrated that the output of the evaluation parameters of the deep learning algorithm has an accuracy rate of 98.66%, a recall rate of 63.47%, and a f-measure rate of 68.99%.

Recommendations for improving the detection and prediction of faults in the air pressure system of heavy trucks are outlined below:

1- Enhanced Sensor Technology: Upgrade sensor technology to capture more detailed and accurate data on the air pressure system's status. This can involve using advanced sensors with higher precision and reliability.
2- Integration of Additional Data Sources: Incorporate data from additional sources such as vehicle usage patterns, environmental conditions, and maintenance records. This comprehensive dataset can provide deeper insights into potential fault patterns.
3- Continuous Monitoring and Real-Time Analysis: Implement real-time monitoring of the air pressure system coupled with continuous data analysis. This approach enables prompt detection of abnormalities and immediate response to potential issues.
4- Utilization of Advanced Machine Learning Techniques: Explore advanced machine learning techniques, such as deep learning, recurrent neural networks (RNNs), or convolutional neural networks (CNNs), to improve fault detection and prediction accuracy. These algorithms can handle complex data patterns more effectively.
5- Feature Engineering and Selection: Conduct thorough feature engineering to extract relevant features from the dataset and select the most informative ones for model training. This process can enhance the model's predictive performance by focusing on key indicators of system health.
6- Ensemble Learning: Implement ensemble learning techniques that combine predictions from multiple models to improve overall performance and robustness. Ensemble methods, such as random forests or gradient boosting, can effectively mitigate individual model biases and errors.
7- Regular Model Updating and Maintenance: Continuously update and maintain the predictive models to adapt to changing operating conditions and evolving fault patterns. Regular retraining with fresh data ensures that the models remain accurate and reliable over time.
8- Collaborative Data Sharing and Analysis: Foster collaboration among different stakeholders in the trucking industry to share anonymized data

and insights. This collective approach can facilitate the identification of common fault patterns and the development of more effective predictive models.

9- Feedback Mechanism Implementation: Establish a feedback mechanism to incorporate real-world outcomes and user feedback into the model refinement process. This iterative approach helps in identifying and addressing model weaknesses and improving overall performance.

10- Deployment of Predictive Maintenance Strategies: Integrate fault prediction models into predictive maintenance strategies to proactively schedule maintenance activities and prevent potential breakdowns. This proactive approach minimizes downtime and enhances overall operational efficiency.

Disclosure statement and funding

The authors declare no potential conflicts of interest. The present study received no financial support from any organization or institution.

Acknowledgment

We would like to give special thanks to all the participants in this study.

References

[1] Selvi, K. T., Praveena, N., Pratheksha, K., Ragunanthan, S., & Thamilselvan, R. (2022, February 23-25). *Air Pressure System Failure Prediction and Classification in Scania Trucks using Machine Learning* [Conference session]. Second International Conference on Artificial Intelligence and Smart Energy. Coimbatore, India. https://doi.org/10.1109/ICAIS53314.2022.9742716

[2] Cerqueira, V., Pinto, F., Sá, C., & Soares, C. (2016). Combining Boosted Trees with Metafeature Engineering for Predictive Maintenance. In H. Boström, A. Knobbe, C. Soares, & P. Papapetrou (Eds.), *Advances in Intelligent Data Analysis XV* (pp. 393-397). Springer International Publishing. https://doi.org/10.1007/978-3-319-46349-0_35

[3] Rawat, S. (2022). Predict component failure related with air pressure system at scania trucks using various machine learning methods. 1-10. https://doi.org/10.13140/RG.2.2.29587.20003

[4] Yazdinejad, A., Bohlooli, A., & Jamshidi, K. (2018). Efficient design and hardware implementation of the OpenFlow v1.3 Switch on the Virtex-6 FPGA ML605. *The Journal of Supercomputing*, *74*(3), 1299-1320. https://doi.org/10.1007/s11227-017-2175-7

[5] Akarte, M. M., & Hemachandra, N. (2018, December 16-19). *Predictive maintenance of air pressure system using boosting trees: A machine learning approach* [Conference session]. 51st Annual Convention of Operational Research Society of India & International Conference, Mumbai, India. https://mrugankakarte.github.io/files/ORSI2018.pdf

[6] Costa, C. F., & Nascimento, M. A. (2016). IDA 2016 Industrial Challenge: Using Machine Learning for Predicting Failures. In H. Boström, A. Knobbe, C. Soares, & P. Papapetrou

(Eds.), *Advances in Intelligent Data Analysis XV* (pp. 381-386). Springer International Publishing. https://doi.org/10.1007/978-3-319-46349-0_33

[7] Costa, L. M. D. (2019). *Voting Nearest Neighbors: SVM Constraints Selection Algorithm Based on K-Nearest Neighbors* [Doctor, University of Rhode Island]. South Kingstown, Rhode Island. https://www.proquest.com/openview/5c27a7f79ad0ddae70e709d9236e0d07/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y#page=133

[8] Gondek, C., Hafner, D., & Sampson, O. R. (2016). Prediction of Failures in the Air Pressure System of Scania Trucks Using a Random Forest and Feature Engineering. In H. Boström, A. Knobbe, C. Soares, & P. Papapetrou (Eds.), *Advances in Intelligent Data Analysis XV* (pp. 398-402). Springer International Publishing. https://doi.org/10.1007/978-3-319-46349-0_36

[9] Lokesh, Y., Nikhil, K. S. S., Kumar, E. V., & Mohan, B. G. K. (2020, May 13-15). *Truck APS Failure Detection using Machine Learning* [Conference session]. 4th International Conference on Intelligent Computing and Control Systems, Madurai, India. https://doi.org/10.1109/ICICCS48265.2020.9121019

[10] Ozan, E. C., Riabchenko, E., Kiranyaz, S., & Gabbouj, M. (2016). An Optimized k-NN Approach for Classification on Imbalanced Datasets with Missing Data. In H. Boström, A. Knobbe, C. Soares, & P. Papapetrou (Eds.), *Advances in Intelligent Data Analysis XV* (pp. 387-392). Springer International Publishing. https://doi.org/10.1007/978-3-319-46349-0_34

[11] Rafsunjani, S., Safa, R. S., Al Imran, A., Rahim, M. S., & Nandi, D. (2019). An empirical comparison of missing value imputation techniques on APS failure prediction. *International Journal of Information Technology and Computer Science*, *11*(2), 21-29. https://doi.org/10.5815/ijitcs.2019.02.0

[12] Syed, M. N., Hassan, M. R., Ahmad, I., Hassan, M. M., & Albuquerque, V. H. C. D. (2021). A Novel Linear Classifier for Class Imbalance Data Arising in Failure-Prone Air Pressure Systems. *Institute of Electrical and Electronics Engineers Access*, *9*, 4211-4222. https://doi.org/10.1109/ACCESS.2020.3047790

[13] Nakhodchi, S., Zolfaghari, B., Yazdinejad, A., & Dehghantanha, A. (2021, December 13-15). *SteelEye: An Application-Layer Attack Detection and Attribution Model in Industrial Control Systems using Semi-Deep Learning* [Conference session]. 18th International Conference on Privacy, Security and Trust, Auckland, New Zealand. https://doi.org/10.1109/PST52912.2021.9647777

[14] Rabieinejad, E., Yazdinejad, A., Parizi, R. M., & Dehghantanha, A. (2023). Generative adversarial networks for cyber threat hunting in ethereum blockchain. *Distributed Ledger Technologies: Research and Practice*, *2*(2), 1-19. https://doi.org/10.1145/3584666

[15] Yazdinejad, A., Parizi, R. M., Srivastava, G., & Dehghantanha, A. (2020, December 7-11). *Making Sense of Blockchain for AI Deepfakes Technology* [Conference session]. 2020 Institute of Electrical and Electronics Engineers Globecom Workshops, Taipei, Taiwan. https://doi.org/10.1109/GCWkshps50303.2020.9367545

[16] Yazdinejad, A., Dehghantanha, A., Parizi, R. M., & Epiphaniou, G. (2023). An optimized fuzzy deep learning model for data classification based on NSGA-II. *Neurocomputing*, *522*(4), 116-128. https://doi.org/10.1016/j.neucom.2022.12.027

[17] Sisk, M., Majlis, M., Page, C., & Yazdinejad, A. (2023). Analyzing xai metrics: Summary of the literature review. *Authorea Preprints*, 1-6. https://doi.org/10.36227/techrxiv.21262041.v1

[18] Yazdinejad, A., Dehghantanha, A., Parizi, R. M., Hammoudeh, M., Karimipour, H., & Srivastava, G. (2022). Block Hunter: Federated Learning for Cyber Threat Hunting in Blockchain-Based IIoT Networks. *Institute of Electrical and Electronics Engineers Transactions on Industrial Informatics*, *18*(11), 8356-8366. https://doi.org/10.1109/TII.2022.3168011

[19] Yazdinejad, A., Parizi, R. M., Dehghantanha, A., Srivastava, G., Mohan, S., & Rababah, A. M. (2020). Cost optimization of secure routing with untrusted devices in software defined networking. *Journal of Parallel and Distributed Computing*, *143*, 36-46. https://doi.org/10.1016/j.jpdc.2020.03.021

[20] Scania. (2017). *APS Failure at Scania Trucks* [Data set]. University of California Irvine Machine Learning Repository. https://doi.org/10.24432/C51S51

[21] Uska, M. Z., Wirasasmita, R. H., Usuluddin, U., & Arianti, B. D. D. (2020). Evaluation of rapidminer-aplication in data mining learning using persiva model. *Edumatic: Jurnal Pendidikan Informatika*, *4*(2), 164-171. https://doi.org/10.29408/edumatic.v4i2.2688

[22] Yazdinejad, A., Parizi, R. M., Dehghantanha, A., Zhang, Q., & Choo, K. K. R. (2020). An Energy-Efficient SDN Controller Architecture for IoT Networks With Blockchain-Based Security. *Institute of Electrical and Electronics Engineers Transactions on Services Computing*, *13*(4), 625-638. https://doi.org/10.1109/TSC.2020.2966970

[23] Yazdinejad, A., Dehghantanha, A., Parizi, R. M., Srivastava, G., & Karimipour, H. (2023). Secure Intelligent Fuzzy Blockchain Framework: Effective Threat Detection in IoT Networks. *Computers in Industry*, *144*(4), 103801. https://doi.org/10.1016/j.compind.2022.103801

[24] Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, *81*(3), 429-450. https://doi.org/10.1007/s10472-017-9564-8