Minimizing the Repair Cost of the Air Pressure System of Scania Trucks Using a Deep Learning Algorithm

Kazem Taghandiki 1 and Morteza DallakehNejad 2

1Department of Computer Engineering
2Affiliation not available

October 30, 2023

Abstract

Air pressure systems play an essential role in Scania trucks, so that the correct operation of the braking system and gear shifting system of Scania cars depends on the health of the air pressure system. The presence of the central control unit (ECU) in the air pressure system has caused the collection of different information about its different status, which can be stored and checked in the form of different data sets.
Minimizing the repair cost of the air pressure system of Scania trucks using a deep learning algorithm

Kazem Taghandiki*, Morteza DallakehNejad†

*Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran  
ktaghandiki@tvu.ac.ir
† Department of Mechanical Engineering, Technical and Vocational University (TVU), Tehran, Iran  
mdallakehnejad@tvu.ac.ir

Abstract—Air pressure systems play an essential role in Scania trucks, so that the correct operation of the braking system and gear shifting system of Scania cars depends on the health of the air pressure system. The presence of the central control unit (ECU) in the air pressure system has caused the collection of different information about its different status, which can be stored and checked in the form of different data sets. The use of machine learning algorithms in order to detect malfunctions in the air pressure system prevents material and time costs and manual checks in different time frames. A lot of work has been done to diagnose air pressure system failure through data set collected from sensors of various components of Scania trucks with traditional machine learning algorithms such as Decision Tree, SVM, Random Forest, KNN. But since the number of records collected in different intervals by the electronic control unit (ECU) are very large, it is possible to increase the accuracy and speed in diagnosing the failure of the air pressure system of Scania trucks using new machine learning algorithms such as deep learning. also used In this article, feature selection and deep learning algorithms have been used together with the Than activation function to detect and predict failure in the air pressure system of heavy trucks, as well as a comparison table between the results obtained from the deep learning algorithm and Five traditional methods of machine learning have been developed. The results and observations show the accuracy (98.66%) in the output obtained from the deep learning algorithm.

Index Terms—Deep learning, air pressure system, preprocessing, feature selection, dataset

1 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 air pressure system for proper operation of the brake and gear shifting system [1]. Since Scania trucks are used in many industries and long journeys, monitoring the air pressure system of Scania trucks is very important 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 example, 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 works [5], [6], [7], [8], [9], [10], [11], [12] have been done to identify and quickly diagnose the failure of the air pressure system in Scania trucks. But these works are mostly focused on using traditional machine learning algorithms such as decision tree, logistic regression, SVM, KNN, Random Forest [13], [14], [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, the APS dataset has data collected from ECUs in Scania trucks and has two training and testing sections with a record number of 60,000 and 16,000[9]. In the proposed approach, first, by using 3 feature selection algorithms, 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. But since many of the selected features have noise values, in the next step, using the Replace All Missing operator, the process of repairing the noise feature was done automatically. Then, using the Deep Learning algorithm along with the Than activation function, the process of training the deep learning model was carried out through the pre-processed data set [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 2nd part of the article, the work done 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 have been examined. Finally, the conclusion of the proposed approach is discussed in section 5.

2 Related Work

There has been 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] have 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 have balanced the APS dataset. Then, using traditional machine learning algorithms such as SVM, Logistic Regression, Random Forest, KNN, SGD, Decision Tree and Naive Bayes, they have identified and predicted failure in the APS balanced data set. The approach presented by them has the best accuracy in Random Forest and KNN algorithms by 98%.

Cerqueira et al. [2] have 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 have used SMOTE algorithm to balance the number of data set records to solve the problem of APS data set not being balanced, and also used 3 noise feature analysis algorithms to remove features with noisy values. Finally, they used XGBoost and Random Forest algorithm to investigate the cost reduction in the air pressure system of heavy trucks, and XGBoost algorithm showed the best performance in the output results.

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

Costa et al. [5] 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 have predicted the failure of parts related to the air pressure system in heavy trucks, their results are the most accuracy (92.5%) for the Random Forest algorithm.

Rafsanjani et al [6] 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 has been analyzed using 5 machine learning algorithms Naive Bayes, SVM, KNN, Random Forest and Gradient Boosted Tree in order to identify and diagnose the failure of the air pressure system. Their results show the highest accuracy (98%) for the Random Forest algorithm.

3 The Proposed Approach

The proposed system consists of 4 steps, Figure 1 shows the implementation process of the proposed approach. Next, each of the steps in Figure 1 are explained in detail.

3.1 Select dataset

In this research, the APS dataset available on the UCI website has been 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 also 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.

In the 170 specified features, it means the number of input features and the tag feature (class) is not included in this number. 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.

3.2 Feature selection

In this step, selection and weighting of important features from the APS data set was done separately with 3 feature
selection algorithms, information gain, correlation and svm. The working method was that a separate list was prepared for each of the feature selection algorithms, finally the important and common features in the output of each of these 3 algorithms (3 lists produced) as important features and have More weight was chosen. Therefore, among the 170 features of the APS educational data set, 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 RapidMiner software [21].

3.3 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, therefore, it is necessary to take measures to resolve these types of inconsistencies, which It 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 data. Therefore, at this stage, 4 pre-processing processes mentioned below have been done [22].

1. 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, therefore, by using the Nominal to Binominal operator, the data type of this attribute is changed from polynomial to binomial.

2. Specify label attribute
Due to the fact that when adding the training and testing data set, the label (class) feature must be recognized, therefore, the Set Role operator has been used to specify the class feature as the label (class) feature.

3. 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.

4. Identification and quantification of noise features
Because most of the features of the APS dataset have noise values. Therefore, by examining 21 input features, it was determined that all of them have limited noise values. Therefore, by using the Replace All Missing operator, the average values in each of the numerical features of the data set 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.

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

4 DEEP LEARNING MODELING
In the third phase of the proposed approach, a deep learning model is built from the pre-processed training data set using the deep learning operator and Than activation function. The deep learning model has 4 layers (one input layer, two hidden layers, and one output layer), 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.

Now, it should be checked 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 is received as an input, and the process of prediction and evaluation of the obtained deep learning model is performed on the test data set (with 16,000 records). 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 6 shows the complete process of implementing the deep learning model in the RapidMiner machine learning software.

Finally, using the Performance operator, the results obtained by the machine learning model for the records of the experimental data set have been evaluated, which is fully explained in Section 4.
5 Observations and results

According to the labels predicted by the trained deep learning model and the actual labels of each record of the test data set, the clutter matrix [24] can be created. Now 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 data set.

Now it is possible to evaluate the final model of deep learning by specifying the parameters of the Confusion matrix. Figure 8 shows the results of 5 evaluation criteria of the final deep learning model.

Also, 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.
Fig. 6. Weights assigned to input features in deep learning model

Fig. 7. Label prediction of test dataset records by trained deep learning model

Fig. 8. The complete process of implementing deep learning model in RapidMiner machine learning software

### TABLE 3
Evaluation parameters of traditional machine learning algorithms

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>Accuracy</th>
<th>Error rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Run Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>98.18</td>
<td>1.82</td>
<td>81.34</td>
<td>29.07</td>
<td>42.83</td>
<td>300</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>96.90</td>
<td>3.10</td>
<td>42.11</td>
<td>86.13</td>
<td>56.57</td>
<td>25</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>97.95</td>
<td>2.05</td>
<td>83.10</td>
<td>15.73</td>
<td>28.46</td>
<td>22</td>
</tr>
<tr>
<td>Random Forest</td>
<td>98</td>
<td>1.00</td>
<td>86.36</td>
<td>17.35</td>
<td>28.50</td>
<td>100</td>
</tr>
<tr>
<td>KNN</td>
<td>98.33</td>
<td>1.67</td>
<td>77.34</td>
<td>44.53</td>
<td>56.34</td>
<td>250</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>98.66</td>
<td>1.34</td>
<td>83.46</td>
<td>63.39</td>
<td>68.99</td>
<td>41</td>
</tr>
</tbody>
</table>

6 Conclusion

Air pressure systems play an essential role in Scania trucks, so that the optimal performance of the braking and gear shifting systems of Scania cars depends on the health of the air pressure system. The presence of the sensor in the air pressure system has caused the collection of various information about its condition, which can be stored and checked in the form of various data sets. The use of machine learning algorithms to detect malfunctions in the air pressure system prevents manual checks in different time intervals. In this article, deep learning algorithm along with Than activation function and feature selection are used to detect and predict failure in the air pressure system of heavy trucks.
The results and observations from the output of the deep learning algorithm show the accuracy rate (98.66%).

REFERENCES


