

# Interactive feature extraction for diagnostic trouble codes in predictive maintenance

A case study from automotive domain

Parivash Pirasteh\*  
Center for Applied Intelligent  
Systems Research, Halmstad  
University  
Sweden  
parivash.pirasteh@hh.se

Sławomir Nowaczyk  
Center for Applied Intelligent  
Systems Research, Halmstad  
University  
Sweden  
slawomir.nowaczyk@hh.se

Sepideh Pashami  
Center for Applied Intelligent  
Systems Research, Halmstad  
University  
Sweden  
sepideh.pashami@hh.se

Magnus Löwenadler  
Aftermarket Solutions Department,  
Volvo Trucks  
Gothenburg, Sweden  
magnus.lowenadler@volvo.com

Klas Thunberg  
Service Market Products, Volvo Buses  
Gothenburg, Sweden  
Klas.Thunberg@volvo.com

Henrik Ydreskog  
Aftermarket Solutions Department,  
Volvo Trucks  
Gothenburg, Sweden  
henrik.ydreskog@volvo.com

Peter Berck  
Center for Applied Intelligent  
Systems Research, Halmstad  
University  
Sweden  
peter.berck@hh.se

## ABSTRACT

Predicting future maintenance needs of equipment can be addressed in a variety of ways. Methods based on machine learning approaches provide an interesting platform for mining large data sets to find patterns that might correlate with a given fault. In this paper, we approach predictive maintenance as a classification problem and use Random Forest to separate data readouts within a particular time window into those corresponding to faulty and non-faulty component categories. We utilize diagnostic trouble codes (DTCs) as an example of event-based data, and propose four categories of features that can be derived from DTCs as a predictive maintenance framework. We test the approach using large-scale data from a fleet of heavy duty trucks, and show that DTCs can be used within our framework as indicators of imminent failures in different components.

\*Dr. Pirasteh is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

WIDM'19, February 15, 2019, Melbourne, VIC, Australia.

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6296-2/19/02...\$15.00

<https://doi.org/10.1145/3304079.3310288>

## CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by classification; Feature selection;** • **Hardware** → **Failure prediction.**

## KEYWORDS

Predictive maintenance, failure detection, diagnostic trouble codes, feature extraction

## ACM Reference Format:

Parivash Pirasteh, Sławomir Nowaczyk, Sepideh Pashami, Magnus Löwenadler, Klas Thunberg, Henrik Ydreskog, and Peter Berck. 2019. Interactive feature extraction for diagnostic trouble codes in predictive maintenance: A case study from automotive domain. In *Proceedings of WIDM'19*. ACM, New York, NY, USA, Article 4, 10 pages. <https://doi.org/10.1145/3304079.3310288>

## 1 INTRODUCTION

A significant part of the total operating cost in many industries is maintenance of assets and machines. Therefore, being able to predict failure occurrences at an early stage plays an important role in keeping high availability and reliability levels, product quality, and safety requirements. As many products become more and more complex, with increased dependencies among components and decreased development and testing cycles, the challenges related to understanding product usage and modeling life cycles become, in many industries, a major issue. With digitization and connectivity revolution though, comes the increased availability of data from different sources and the increased capability to process this data with more powerful hardware and more complex analysis methods.

Machine learning is one of the most promising approaches for addressing the challenges that arise in predictive maintenance. It has been shown that machine learning, given sufficient amount and quality of the data, can lead to very accurate failure prediction and discovering patterns of interest from sensor measurements [5, 15–17]. The main reason for those successes is the capability of algorithms such as artificial neural networks, random forests or support vector machines to discover patterns that are too complex for human experts to capture. However, those methods are still heavily dependent on careful data cleaning and feature extraction, which makes it especially important to design interactive solutions where experts share their knowledge with a systems in a continuous manner and can provide direct feedback.

Predictive maintenance in the automotive domain is more challenging than it may seem, since vehicles are moving machines and often operate in areas with low network coverage or traveling between countries. This makes continuous monitoring of vehicles difficult, due to for example the cost of wireless communication and the limitations of computational power [10, 11].

Various machine learning methods have been used to identify faulty components, measure the wear and predict failures [5]. However, existing diagnostics models still suffer from a high false positive rate, while also missing detection too often. The main reason is the complexity of the vehicles and their associated environment. Other challenges include imbalanced datasets, where negative samples are dominant, mislabeled samples where the failure has resulted from improper used rather than quality issues, and redundant and noisy data are some factors that would affect the performance of a machine learning method.

DTCs are the example of useful information that captures a lot of engineering expertise that went into designing a complex product. Currently, DTCs are mostly analyzed in a workshop, by technicians trying to determine the underlying cause of the problem that has already manifested and rendered the machine unusable. Thus, the benefits of DTCs are limited to diagnosing a failure, and identifying the associated component. However, in many of today's products, the amount and frequency of DTCs are very large, and they can be continuously analyzed and treated as symptoms of imminent faults. Intuitively, one would expect the increasing in number of DTCs as a sign of anomalous behavior that is developing, even before actual failure.

While DTC data are expected to contain informative events, there is very little work on investigating the possible association between DTCs and failures of different components. One reason is that there are a number of challenges inherent in this kind of data: irregular occurrences, sparsity, complex structure, and usage dependent behavior. This paper proposes a general approach towards using DTCs to create predictive models for maintenance. Our method is based on several different types of features that can be extracted from raw DTC data. The contribution of this paper lies in the empirical analysis and results that demonstrate the potential benefit of incorporating DTCs into the prediction models. Different features and sets of features can be interactively suggested by an expert, evaluated immediately, and the corresponding increase or decrease of accuracy can be presented. Through a case study using large-scale data from a fleet of heavy duty Volvo Trucks, we showcase the difficulties and benefits of applying DTCs as a predictor of

a component failure and show the approach taken for each step in the model.

We formulate the problem of predictive maintenance as a binary classification task, where the positive and negative outcomes are interpreted as the component going to fail or not within a particular time horizon, respectively. In order to explore suitable subset of features so that they are potentially useful in classification, four different categories of features in two levels of individual and aggregated levels are defined. The paper investigates the performance of the classifier by tuning different hyper parameters of the algorithm as well as using different input parameters (size of windows).

Experimental results confirm the feasibility of proposed features and show that the proposed model predict the failure in components using DTC data with a reasonable accuracy. Therefore, if the model predicts that a component is going to fail, then the failure can be prevented in time and successfully treated by predictive maintenance.

This paper is organized as follows: the next section, section 2 reviews the literature on predictive maintenance and feature extraction techniques. We provide detailed characteristics of DTC data in section 3, our overall framework for prediction and feature extraction algorithms are introduced in section 4, and the comprehensive evaluation results in section 5. We discuss future work and conclude the paper in section 6.

## 2 RELATED WORK

Much of the current research in maintenance modeling is concerned with condition-based monitoring which monitors a parameter's condition in a vehicle in order to detect a significant change. Such models are useful for non-complex systems aiming to detect the malfunctioning of an individual sensor. [6] proposed a least squares support vector machine (LS-SVM) framework for optimization of maintenance strategy relying on real-time condition monitoring of industrial machines. Non stationary behaviors of measurement signals is one of the critical tasks in condition monitoring [7]. [8] have treated the failure vibration patterns as novelty detection task. The basic concept of novelty detection is to examine a system's dynamic response to determine if the system significantly deviates from an initial baseline condition [18]. [3] developed a flexible and generic system for novelty detection aiming to diagnose faults when they have already occurred and to have the capability to provide warnings when a fault is still at its early stages. [9] recently provided an extended review on machine prognostics in the context of condition based monitoring. The authors also indicated the need for historical data that include many failures, even "catastrophic" failures for the development of a model for representing system behavior.

In order to monitor the conditions of the operating equipments and provide timely maintenance, [13] suggested an information system to predict failures of medical equipment based on log data. Their suggested approach was developed with an active involvement from domain experts and was evaluated by both machine learning and domain standards. [19] compares a period based approach and event driven based approach to show which prediction approach provides better accuracy and is more suitable for practical use in real world problems. They suggested that period-based

## Interactive feature extraction for DTCs

approach is less sensitive to size of observation window and is suitable for long-term prediction. In this paper we apply a period based approach which consists of an observation window and a prediction window.

Although the use of component logs is common practice in forecasting failure, this data rarely contain explicit information for failure prediction. Expert knowledge has been proven to be of great help in many circumstances and very valuable in building predictive maintenance models [1]. [20] suggested a rule based model for the failure prognostic by taking advantage of the expert knowledge and the analysis of the failure mechanism. [2] reported successful diagnosis applications based on knowledge processing and tracing the complex relations between various signals and possible results in a way that experts do.

Our paper extends a recent work by Wang et al. [14] who proposed a parameterized solution for event/log based analysis. In that approach, four categories of features are extracted and then feature selection is used to identify the most important features for model construction. Our work differs in several aspects. First, certain parts of our work are reliant on domain knowledge, which are not covered in their work. In this way, we extract specific features from the DTCs which rely on domain knowledge. second, we designed new features specifically targeting the variation of input data in the observation window.

**Table 1: An example of a DTC (RVB\_B101320) that consists of four parts.**

RVB	B	1013	20
Family node	System type	Failure component of system	Failure type

### 3 OVERVIEW OF DTCs

Diagnostic trouble codes (also known as fault codes) are the codes that a monitoring system outputs when it detects a malfunction. Each diagnostic trouble code name consists of a node family identifier, followed by one letter indicating the system in relation with diagnostic trouble code. Body, chassis, power train, network and vehicle integration are four systems identifies the "main system" where the fault occurred. The next four digits indicate the failure component of the system and the last two digits of DTC define the failure type. The DTC failure type consists of 16 different failure categories, where each category is associated with 16 subtype failures known as symptoms.

DTCs are indicators of a wide variety of problems, including serious mechanical or electrical issues, or unimportant problems. The simplest variation of these codes are set when a sensor reading is outside of normal range. Primarily DTCs were implemented in vehicles to facilitate maintenance and reparations by providing help to mechanics in workshops [4].

DTCs do not occur at regular time intervals: one DTC might occur frequently, while another one might occur very rarely or never at all. The unique characteristics of DTCs, including the

non regular intervals between the occurrences, the sparsity of the occurrences and their complex structures are some of the challenges met when developing a DTC based predictive maintenance system. Beside this, DTCs can be triggered by a wide range of stimuli and situations. One DTC might have multiple root causes and a single DTC might have multiple symptoms. These challenges must be addressed in order to construct a model based on DTCs for early detection of a component failure.

DTCs are logged in the vehicle's electrical control units while driving and later loaded into a central database when visiting workshops, enabling information to be further analyzed. This database contains detailed information on DTCs occurrences from different components of vehicle. This source of data makes it possible to apply machine learning techniques aimed at discovery of relationship between DTCs and failure in component. Since DTC data contain history of informative events related to the particular component, They seem to be good predictor of failure for a component. This will hopefully help to find defects in vehicles, making it possible to act before problems strike a larger part of the vehicle population.

Figure 1 shows a subset of DTCs that are sent during the lifetime of multiple trucks.

### 4 METHODS

In this paper we develop a classification-based predictive maintenance framework. The goal is to identify components that are going to fail within a given prediction window, based on DTC data.

The first step towards building the prediction model is merging two datasets: the DTC data and vehicle repair data. The vehicle repair data contains historical information related to repairs that were performed on a component. We use period-based approach as a framework for prediction (see Figure 2). The period-based approach consists of three parts: observation window, lead time and prediction window. The vector of features is extracted from every observation window and use as an input data. The observation window is composed of a set of consecutive time intervals called sub-windows. Each sub-window spans fourteen days, corresponding to the time interval of DTC data collection. In this paper we use random forest models for prediction task, since random forest algorithm works very fast and tends to perform well with complex input data. We set a maximum tree depth to 10 and the number of tree sets included in each stage is equal to 100.

Working with DTC, there are some common parts of the knowledge which can be extracted, but we can gain a lot of new knowledge and distinctive features from experts to develop a framework for predictive maintenance.

Lead time is the minimum time interval that we would like to have preceding the failure time. A long enough lead time would provide sufficient time to deal with the failure that might arise. Since the goal of this paper is limited to outline the relation between DTCs and component failure rather than to create a generic model, for simplicity we consider lead time equal to zero. The period based framework predicts the future failure over a certain period, called prediction window. The class label of an instance is defined positive if there is any component failure inside the prediction window, otherwise is set to negative.

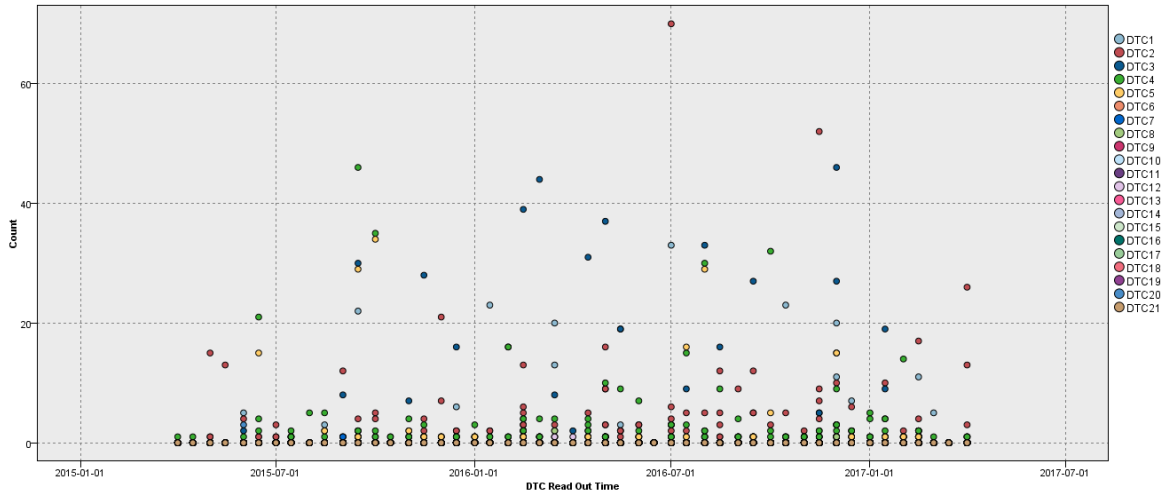


Figure 1: An example of DTC occurrences during two years period of multiple trucks.

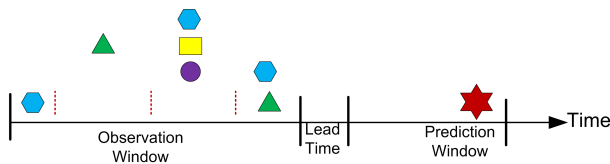


Figure 2: Period based framework consists of three parts, observation window, lead time and prediction window.

We start from truck’s delivery date to create instances and then move the prediction point along the lifetime of the truck by a constant parameter as moving step. By repeating this process for all vehicles, we can obtain all the training and testing instances to build the model. The total number of learning instances would be dependent on the size of the windows generated from historical data and moving step. Our goal is to perform a thorough study on factors such as the size of observation and prediction windows and choices of features that could impact the performance of the classifier.

### 4.1 Level Definition of DTCs

As explained in section 3, DTCs can decompose to different parts. In order to monitor changes in smaller parts, we propose to add an extra level of definition in DTCs. We apply the suggested feature extraction techniques on two level of DTCs, namely: individual level and aggregated level.

**4.1.1 Individual Level.** The natural solution is to treat all available DTCs as independent observations. In principle, keeping the individual identity of each DTC allows the classifier to exploit any truly informative relations and patterns among them. Such treatment of DTCs is very close to what experts and technicians do in workshops today – they try to find any possible connection between an observed DTC and failed components. It is also easier to interpret

the results at this level, since an expert will be able to estimate the importance of individual DTC in connection with component failure.

In practice, however, the amount of noise and the complexity of the data are expected to make the prediction task prohibitively difficult. For instance if different failure types convey different messages related to component failure then focusing on individual level is inadequate to prediction task.

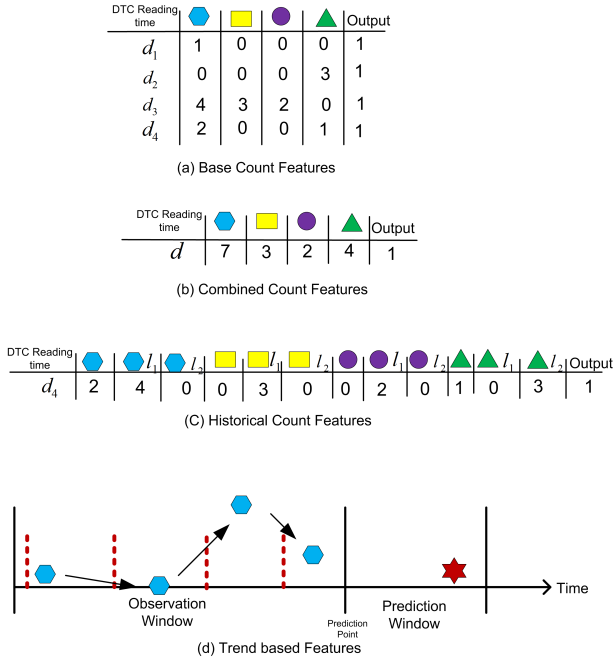
**4.1.2 Aggregated Level.** DTCs have complex structure and are correspond to different symptoms of problems attached to their failure node, system type, failure type and failure component. In aggregated level of DTCs, each DTC decomposes into four units where each unit represents a message from DTC. So instead of tracking single DTC, we aim to monitor changes in each unit of DTCs. For example, if changes in system type of DTC has a relation with component failure then looking in aggregated level of DTCs may be more helpful than individual level. Capturing DTCs at aggregated levels helps reveal systematic deviations in DTCs that may be difficult to detect from individual perspective.

The idea of using aggregated level came from expert knowledge since DTCs are designed to monitor certain functions of different components in the vehicles. Aggregation of DTCs based on functional area could generate new information for detecting the malfunction of the components.

### 4.2 Feature Extraction

Feature extraction is process of extracting some valuable parameters for further processing of the input data [12]. The goal of feature extraction is to take relevant and most important information for classification from raw data. Being highly distinctive, being easy to extract and invariant to noise are properties that we would like to have in proposed features. In classification, it is important to extract features that are focused on discriminating between classes

## Interactive feature extraction for DTCs



**Figure 3: Different features that are extracted from observation window.**

and choosing the right subset of features to get the best generalization. The proposed methodology is presented for extracting the information from experts to develop a framework for predictive maintenance.

After generating all instances from historical data, we extract features from the observation window. We define four types of features: Base count features, combined count features, historical count features, and trend based feature.

- **Base Count Features.** The first category of features represents the number of DTCs that triggered within each individual sub-window. We denote the base counts feature in individual level as  $BCF_i = \{c_i, i \in [1, N]\}$ , where  $N$  is the number of DTC types. Thus the length of  $BCF$  in individual level is  $N$ . In aggregated level, we define four groups of features which represent family node, system type, failure type and failure component of system of a DTC. The definition of  $BCF$  in this level includes the number of times that each part of DTC occurred per sub-window as  $BCF_{FN} = \{c_i, i \in \{FN\}\}$ ,  $BCF_{FT} = \{c_i, i \in \{FT\}\}$ ,  $BCF_{ST} = \{c_i, i \in \{ST\}\}$  and  $BCF_{FC} = \{c_i, i \in \{FC\}\}$ . Thus the length of  $BCF$  in aggregated level is  $(FN + FT + ST + FC)$ . Where  $FN$  is the number of different groups in failure node,  $FT$  is the number of different failure types,  $ST$  is the number of different system types and  $FC$  is number of different failure components. If the number of sub-windows is  $S$ , the size of input matrix is  $S*N$  in individual level and  $S*(FN + FT + ST + FC)$  in aggregated level.
- **Combined Count Features.** In this category of features, the number of occurrences of DTCs in different sub windows are

aggregated into a single value. Thus, for each observation window, there is a value representing the count for each DTC. The number of features in this group is equal to base count features both in individual and aggregated level but the dimensionality of input matrix is reduced  $1/S$ .

- **Historical Count Features.** This category of feature captures DTC occurrences from current and former sub-windows. The dimension of input feature space is  $3*N$  in individual level and  $3*(FN + FT + ST + FC)$  in aggregated level. This growth in number of features would triple the size of input matrix as well.
- **Trend-based Features.** This category of features attempts to capture the variation in DTCs between two consecutive sub-windows. We denote the trend based feature as  $T = \{(c_{i,j} - c_{i,j-1})/c_{i,j-1}, i \in [1, N], j \in [1, S]\}$  where  $N$  is the number of DTC types and  $S$  is the number of sub-window. In aggregated level, we focus on variation in each of four parts in DTCs. The procedure is same as described in base count features for aggregated level. These features have the same length of features and dimensionality of input matrix as base count features both in individual and aggregated level. The efficiency of this feature is based on the fact that it is sensitive to variation in DTC occurrences.

Figure 3 visualizes the differences between these four feature categories.

## 5 EXPERIMENTS

In the experimental section, we present results from experiments aimed at evaluating the proposed framework. These experiments showcase the effectiveness of extracted features and the idea of applying DTCs in two levels in the learning process. For all experiments the process of data preparation, instance generation and feature extraction have been done based on configuration of experiment. The first group of experiments was devoted to examining the different parameters that could influence in configuration of proposed framework and as a result the performance of prediction models. For instance, by assigning different values to length of observation window and prediction window, different sets of instances will be generated and as a consequence different results are obtained. Moreover, this experiment leads us to compare the usefulness of different extracted features in predicting the failure of component. The second category of experiment has focused on evaluating the effectiveness of combinations of feature classes in prediction.

An air suspension component and a powertrain component have been selected as example components due to their different functionality. Another reasons for picking these two components are high failure rate and their attendant costs. Applying the same set of DTCs as indicator of failures in these two components enables us to explore the possible relation between DTCs and failure in each component and to compare the effectiveness of the models in different parameter settings (e.g. different size of observation window and prediction window) based on accuracy of classification.

The data used in these experiments includes the records of 1.000 individual DTCs across 30.000 Volvo trucks with delivery date from January 2015 to January 2018. In average DTCs are triggered 50

times during lifelong of a truck. We used data from 22500 trucks to train the prediction model, and tested the model with data from the remaining trucks. It is also worth noting that since the set of trucks for training and testing are different, all the instances of same truck would belong to either training or testing, in order to avoid possible over-fitting problem.

### 5.1 Performance comparison on different settings

The first series of experiments concerns the performance of prediction model taking into account different configuration of model parameters. This experiment allows us to investigate how the model reacts to using different levels of DTCs, which feature provides the better discrimination and most importantly, what is the best configuration for the size of observation window and prediction window. For simplicity of parameter tuning, the size of sub-window is fixed to two weeks. The performance was estimated using five-fold cross-validation. The cross-validation process is then repeated on training data five times with each of the five folds used exactly once as validation data. After that, five models have been built based on the divisions of training data and the models were applied to test data. The reported results are based on performance of classifier on test set.

Two target components have been used in these experiments (an air suspension and a powertrain component), since different components have distinctively different degradation times and mechanisms. Other parameters, such as the size of the prediction window (which corresponds to the lead time needed to schedule a workshop visit) also depend on business requirements that vary across different components. Therefore, the same setting of parameters cannot be applied for different components in deployment and it is vital to examine carefully which setting is more appropriate for each component.

Figures 4 and 5, as well as 6 and 7 show the results for different experimental settings: aggregated and individual DTC levels, respectively, for powertrain and air suspension components. The subplots (a-d) demonstrate different categories of extracted features, in particular how area under ROC curve (AUC) changes depending on different settings for prediction and observation windows.

Generally, when using the base count features on DTC data suggests that smaller size of observation window leads to better AUC; this is more pronounced in aggregated level, but also shows in individual level. One interpretation of this pattern is that most relevant DTCs occur close to failure date, therefore enlarging size of observation window will reduce the informativeness of the input data. On the other hand, combined count features give better result with larger observation window. This led to the hypothesis that by applying combined count features, we add more information to the model and summation of values from multiple DTC readings work better than a single reading from DTC. The large jump in AUC for combined count features, is because the number of instances is considerably smaller and the task of prediction is easier for classifier.

According to figure 4(c), defining more features by taking the reading from former sub-windows doesn't improve AUC. The lowest range of AUC values belongs to trend-based features. Figure4(d) suggests that trend-based features cannot play a significant rule

in prediction of failure and there is no pattern between size of observation window and prediction window and accuracy of model applying this feature.

Comparison of results indicates that base count features in aggregated level work slightly better than individual level. Another observation is that the classification model trained with base count features or historical count features have better performance than trend based features. But the improvement in AUC for historical count features compared to based count features, which is provided by adding values from former sub-windows is non-significant.

## 6 CONCLUSIONS AND FUTURE WORK

Unplanned downtime is both dangerous and costly, therefore there is a strong push today to increase the dependability of vehicles by continuously monitoring their health status. The ultimate goal of predictive maintenance is to monitor the system, diagnose the potential problems and forecast the degradation of components to compensate, or maintain them just-in-time.

This paper explores the possible connection between DTCs and a component failure. Our approach was developed based on different features that are extracted from DTCs. Features in use are capturing important aspect of DTCs that are not directly observable in data. Additionally, we defined two levels in DTCs. The individual level is associated with each DTC and aggregated level is associated with different units within DTCs. We used one air suspension and one powertrain component to validate our method. The experimental results confirm the effectiveness of applying DTCs in predicting component failures under certain circumstances. In general, the results demonstrate that applying DTCs in the aggregated level provides better performance than in the individual level. Combined count features in aggregated level of DTCs were shown to outperform other settings for the classification task.

In this paper we present the design, development and outputs of a diagnosis system which employs interactions with expert through-out the diagnosis process. We found that interactive expert system can be used for prediction of failure in a component where expert provides valuable knowledge for building and refining the features.

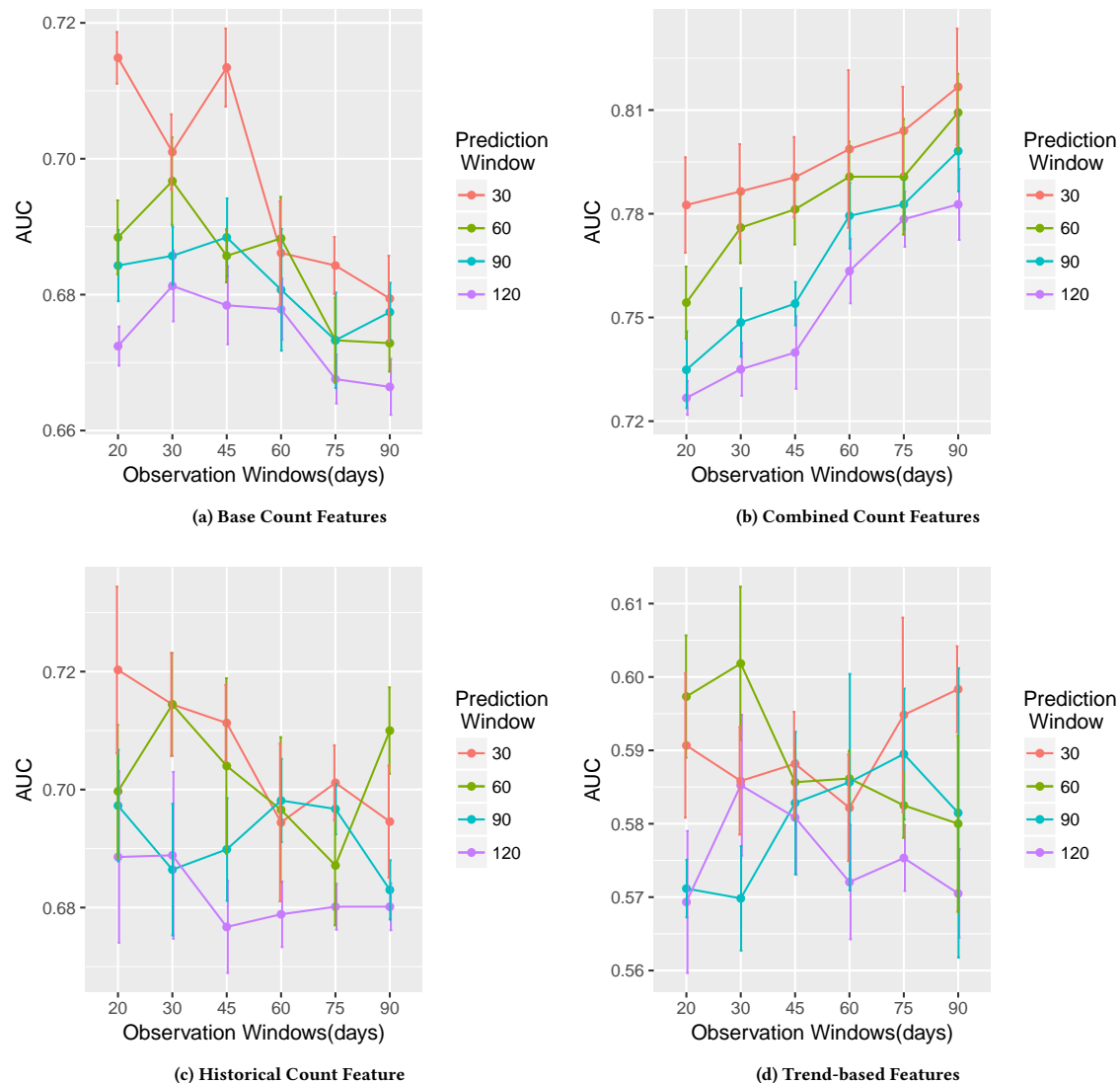
Aside from features defined in paper, other features that may have discriminating power in classifying failures and have not introduced in this paper are features based on sequential patterns and statistics based features.

For new challenges we propose to incorporate domain knowledge in the analysis. The domain knowledge could be expressed in the models in different ways, for example by grouping the repairs into failure modes. So, in one group the repairs might be a result of wear out, while the other could cover repairs associated with mechanical damage. In this way finding possible association between DTCs and failure becomes easier.

In order to improve classification accuracy one can use ensemble methods, where data features vectors are fed into several classifiers and the results are then combined by a voting schema.

A more detailed study in predictive maintenance should be made by integrating all available features (ex. features based on specifications of vehicles and features based on parameters such as fuel consumption) and DTCs features in order to take full advantage of all available information which is not in the scope of this study. The

## Interactive feature extraction for DTCs

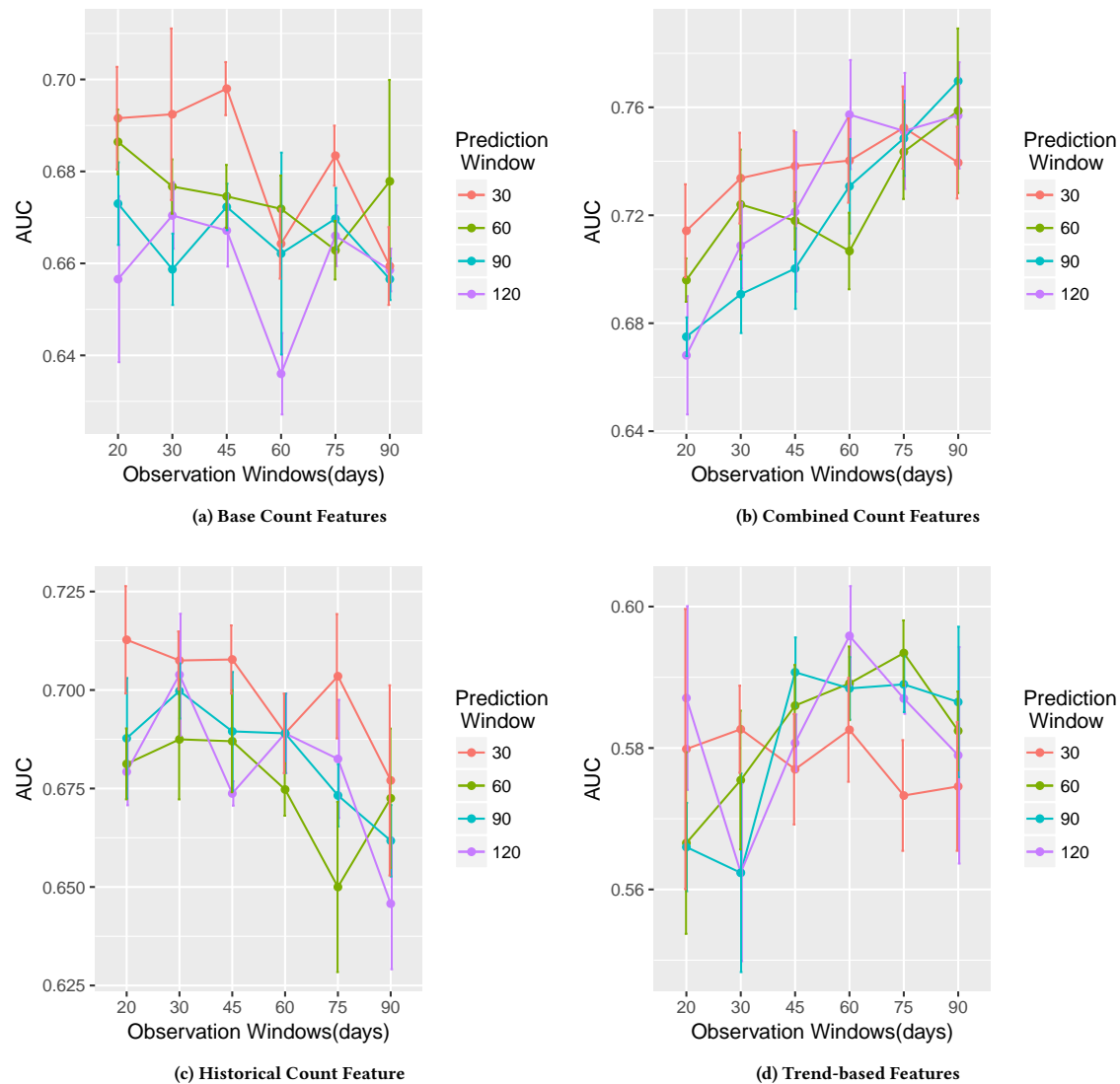


**Figure 4: AUC values with respect to size of observation window and prediction window for powertrain component in aggregated level**

results of this research may lead to profound understanding of different feature combinations and thereby to find the most effective classifications.

## REFERENCES

- [1] Dentcho Batanov, Nagen Nagarur, and Prapan Nitikhunkasem. 1993. EXPERT-MM: A knowledge-based system for maintenance management. *Artificial intelligence in engineering* 8, 4 (1993), 283–291.
- [2] Christine W Chan. 2005. An expert decision support system for monitoring and diagnosis of petroleum production and separation processes. *Expert Systems with Applications* 29, 1 (2005), 131–143.
- [3] Christos Emmanouilidis, Erkki Jantunen, and John MacIntyre. 2006. Flexible software for condition monitoring, incorporating novelty detection and diagnostics. *Computers in industry* 57, 6 (2006), 516–527.
- [4] Moa Fransson and Lisa Fähræus. 2015. Finding Patterns in Vehicle Diagnostic Trouble Codes: A data mining study applying associative classification.
- [5] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing* 20, 7 (2006), 1483–1510.
- [6] Rocco Langone, Carlos Alzate, Bart De Ketelaere, Jonas Vlasselaer, Wannes Meert, and Johan AK Suykens. 2015. LS-SVM based spectral clustering and regression for predicting maintenance of industrial machines. *Engineering Applications of Artificial Intelligence* 37 (2015), 268–278.
- [7] Xiandong Ma. 2011. Novel early warning fault detection for wind-turbine-based DG systems. In *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on*. IEEE, 1–6.
- [8] Jordan McBain and Markus Timusk. 2011. Feature extraction for novelty detection as applied to fault detection in machinery. *Pattern Recognition Letters* 32, 7 (2011), 1054–1061.
- [9] Ying Peng, Ming Dong, and Ming Jian Zuo. 2010. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology* 50, 1 (2010), 297–313.
- [10] Rune Prytz, Slawomir Nowaczyk, Thorsteinn Rognvaldsson, and Stefan Bytner. 2015. Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data. *Engineering applications of artificial intelligence* 41 (2015), 139–150.



**Figure 5: AUC values with respect to size of observation window and prediction window for powertrain component in individual Level**

- [11] Thorsteinn Rögnvaldsson, Slawomir Nowaczyk, Stefan Bytner, Rune Prytz, and Magnus Svensson. 2018. Self-monitoring for maintenance of vehicle fleets. *Data Mining and Knowledge Discovery* 32, 2 (01 Mar 2018), 344–384. <https://doi.org/10.1007/s10618-017-0538-6>
- [12] Firoz Shah, Vimal Krishnan VR, Athulya Jayakumar, et al. 2009. Speaker independent automatic emotion recognition from speech: a comparison of MFCCs and discrete wavelet transforms. In *Advances in Recent Technologies in Communication and Computing, 2009. ARTCom'09. International Conference on*. IEEE, 528–531.
- [13] Ruben Sipos, Dmitriy Fradkin, Fabian Moerchen, and Zhuang Wang. 2014. Log-based predictive maintenance. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 1867–1876.
- [14] J Wang, C Li, S Han, S Sarkar, and X Zhou. 2017. Predictive maintenance based on event-log analysis: A case study. *IBM Journal of Research and Development* 61, 1 (2017), 11–121.
- [15] Achmad Widodo and Bo-Suk Yang. 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing* 21, 6 (2007), 2560–2574.
- [16] Dazhong Wu, Connor Jennings, Janis Terpenney, Robert X Gao, and Soundar Kumara. 2017. A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering* 139, 7 (2017), 071018.
- [17] Sze-jung Wu, Nagi Gebrael, Mark A Lawley, and Yuehwern Yih. 2007. A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37, 2 (2007), 226–236.
- [18] Ai-Min Yan, Pascal De Boe, and Jean-Claude Golinval. 2004. Structural damage diagnosis by Kalman model based on stochastic subspace identification. *Structural Health Monitoring* 3, 2 (2004), 103–119.
- [19] Li Yu, Ziming Zheng, Zhiling Lan, and Susan Coghlan. 2011. Practical online failure prediction for blue gene/p: Period-based vs event-driven. In *Dependable Systems and Networks Workshops (DSN-W), 2011 IEEE/IFIP 41st International Conference on*. IEEE, 259–264.
- [20] Zhi-Jie Zhou, Chang-Hua Hu, Wen-Bin Wang, Bang-Cheng Zhang, Dong-Ling Xu, and Jian-Fei Zheng. 2012. Condition-based maintenance of dynamic systems using online failure prognosis and belief rule base. *Expert Systems with Applications* 39, 6 (2012), 6140–6149.



Interactive feature extraction for DTCs

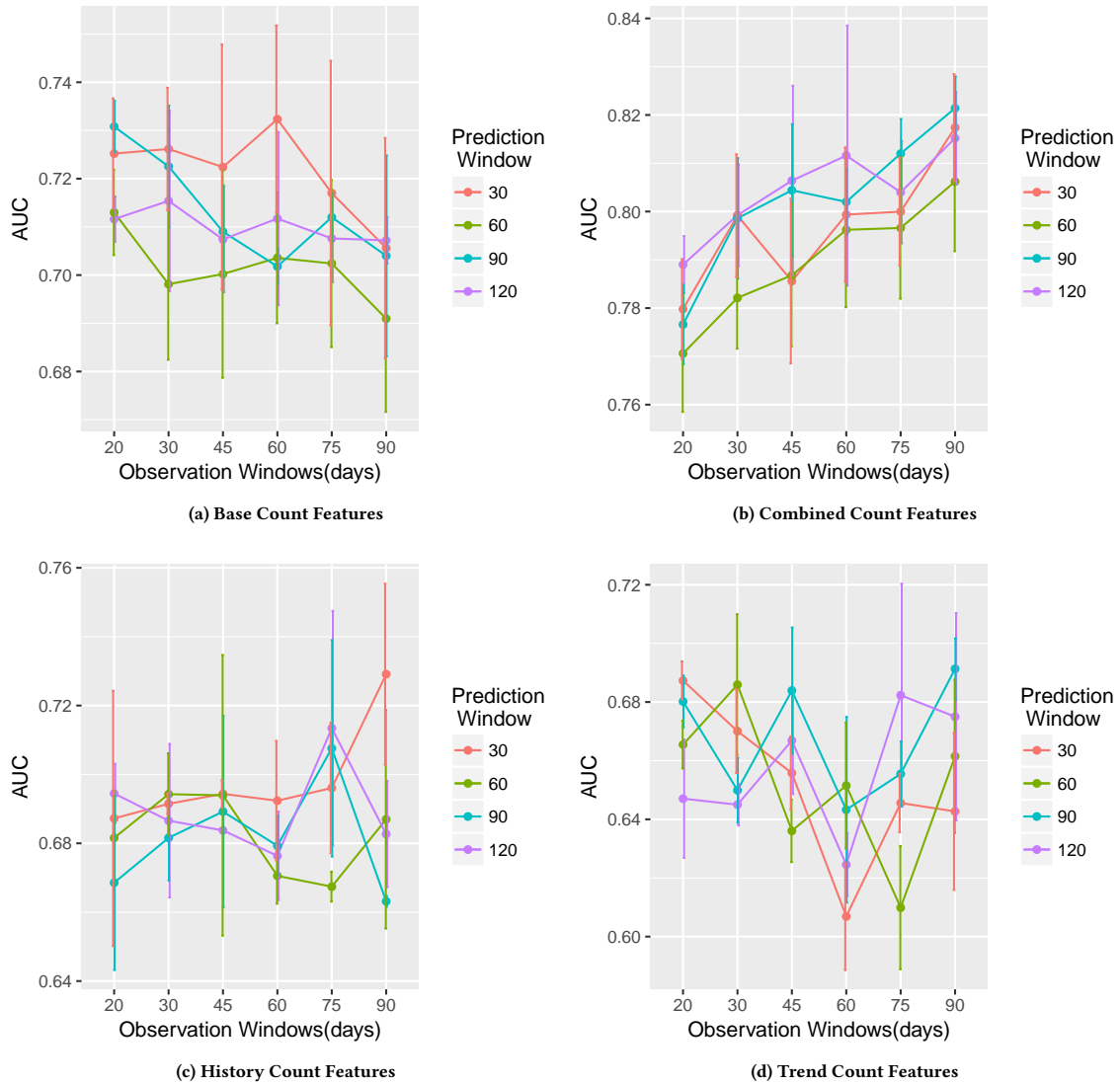
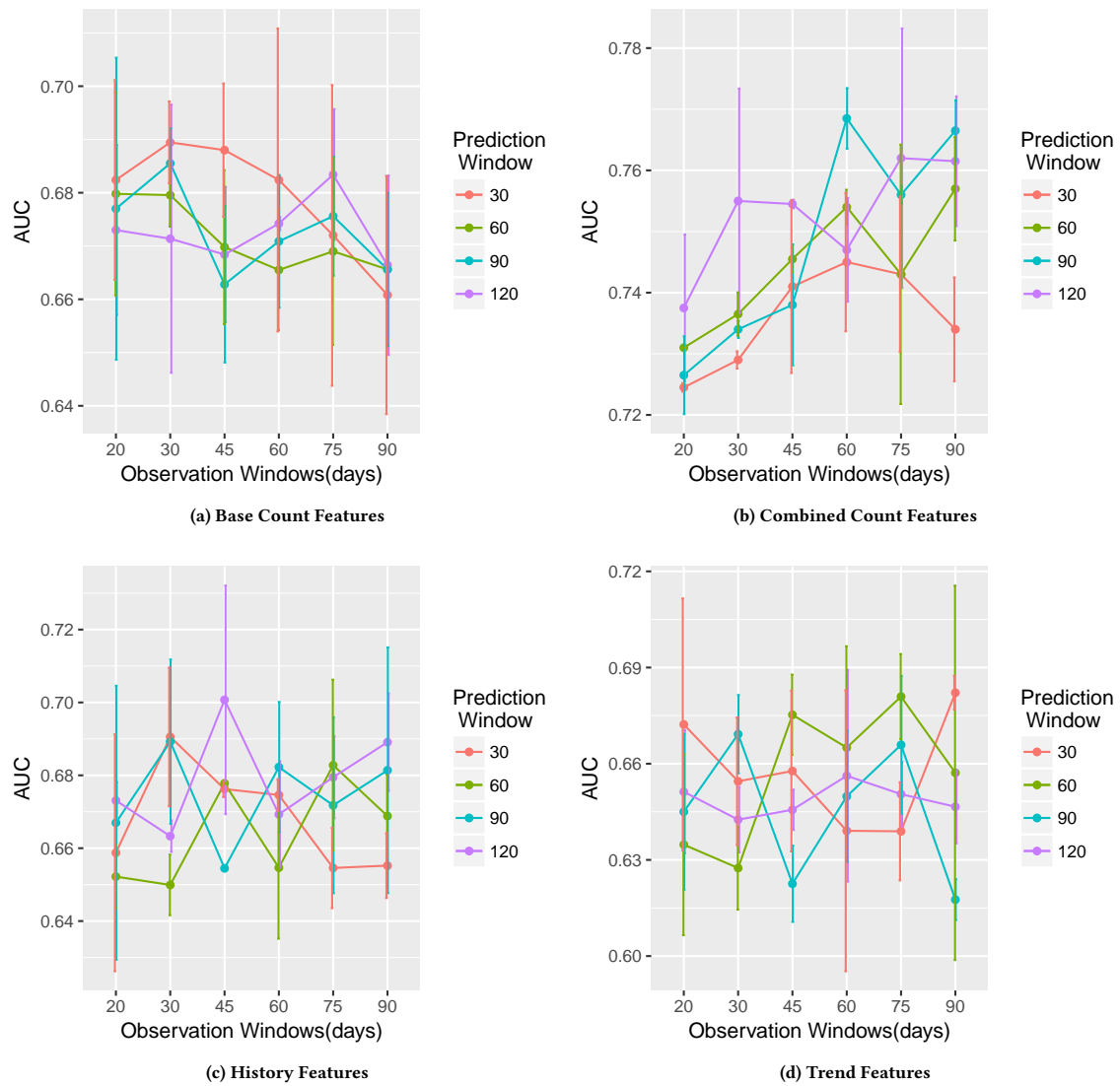


Figure 6: AUC values with respect to size of observation window and prediction window for air suspension component in aggregated level



**Figure 7: AUC values with respect to size of observation window and prediction window for air suspension component in individual levels**