Driving Vehicle Maintenance Decisions using Predictive and Prognostic Maintenance Technology

Diana Perera, Kyle Whirlow, Tim Whalen, PhD, Claire Hughes, James Cooper
Design Interactive, Inc
Orlando, FL
diana.perera@designinteractive.net, kyle.whirlow@designinteractive.net, tim.whalen@designinteractive.net, claire.hughes@designinteractive.net, james.cooper@designinteractive.net

ABSTRACT

In 2022, the U.S. Department of Defense requested a $3.5 billion budget for depot operations and maintenance which includes critical repairs to aircraft, missiles, aircraft carriers, ships, submarines, combat vehicles and other equipment. However, the actual approved budget ($2.1 billion) falls far behind what is needed, threatening the readiness of military vehicles. One way to close this gap is to employ predictive and prognostic maintenance. The U.S. Department of Defense has already prioritized a shift to predictive maintenance solutions as the technology has been proven to optimize standard maintenance scheduling and reduce time, expense, and downtime due to component failure, repairs, and replacements across industries. This approach continues to grow in its usefulness as more powerful machine learning (ML) techniques and greater amounts of data are collected and able to be easily stored. Now that the technology is mature, there is a need to determine which predictive modeling approaches are best suited to military vehicle maintenance.

This paper details research to identify effective approaches for predictive models for vehicle maintenance utilizing a combination of onboard sensor data, historical and peer vehicle data, and maintainer input for predictive maintenance. This includes a comprehensive review of statistical and ML modeling approaches used for vehicle predictive maintenance, with a focus on anomaly detection, fault classification, and remaining useful life estimates from survival regression models. Finally, a case study is presented demonstrating the way these approaches could be applied to U.S. Army vehicle operations and maintenance. Specifically, opportunities and constraints regarding available data and systems are considered, focusing on technical considerations to integrate maintainer feedback into the predictions and decision support recommendations produced by statistical models. This integrated approach can provide more effective and holistic predictive maintenance solutions to help drive vehicle maintenance decisions, reduce costs, and increase readiness in the U.S. military.

ABOUT THE AUTHORS

Diana Perera, M.S. *Diana Perera* is a Research Associate I on the XR Ecosystems Team at DI. Her experience includes conducting human factors research and end-user evaluations of interactive systems and product concepts. She specializes in needs analysis, end-user research, and data analytics. She holds a B.S. in Human Factors Psychology and Master of Science (M.S.) in Human Factors from Embry-Riddle Aeronautical University

Kyle Whirlow, M.S. *Kyle Whirlow* is a Data Scientist at DI. His focus is on development of predictive algorithms and their application. He holds a Master of Science in Data Analytics from the University of Central Florida.

Tim Whalen, Ph.D. *Dr. Whalen* is a Data Scientist and R&D Portfolio Manager at DI. He earned his Ph.D. in Neural Computation from Carnegie Mellon University where he applied statistical modeling, signal processing and ML to analyze how brain waves in Parkinson's disease propagate information and predict motor symptoms. At DI, Tim leads a portfolio of machine perception and information extraction research using a variety of AI/ML methods and has spearheaded DI's development of a modular, generalizable AI perception pipeline.

Claire Hughes, M.S. *Claire Hughes* is a Research Associate III and Project Lead at Design Interactive, Inc. Her focus is on innovative technology delivery to diverse stakeholders, including the Joint Program Committee, and the Army Futures Command. She holds a Master of Science in Human Factors and Systems Engineering from Embry-Riddle Aeronautical University and a Bachelor of Science in Mathematics from Hillsdale College.

James Cooper, M.S. *Lt Colonel (ret.) James Cooper* has over 33 years of military experience in Air Force and Joint Special Operations. Over the course of his military career, James has participated in multiple Joint, and Coalition special operations training exercises and operational missions. Additionally, he has worked with coalition partners to modernize and develop their militaries. This included managing Naval Foreign Military Sales, training, and assets (i.e. frigates, patrol vessels, standard missiles, etc.) at the U.S. Embassy in Manama, Bahrain.

Driving Vehicle Maintenance Decisions using Predictive and Prognostic Maintenance Technology

Diana Perera, Kyle Whirlow, Tim Whalen, PhD, Claire Hughes, James Cooper
Design Interactive, Inc
Orlando, FL
diana.perera@designinteractive.net, kyle.whirlow@designinteractive.net, tim.whalen@designinteractive.net, claire.hughes@designinteractive.net, james.cooper@designinteractive.net

INTRODUCTION

The military needs a maintenance solution that reduces costs and maintains a high level of readiness. One potential solution is utilizing predictive and prognostic maintenance. The U.S. Department of Defense (DoD) has already begun to shift towards the use of predictive maintenance because it has been shown to optimize standard maintenance scheduling and reduce time, expenses, and downtime of equipment. Predictive maintenance is also utilized in a variety of industry sectors, including automobile maintenance and aviation as the costs involving maintenance are a big part of their expenditures (Solanki & Dhall, 2017; Verhagen & De Boer, 2018). The use of predictive maintenance can transform a sector by allowing for more efficiency and ensuring systems are ready when they are needed, as well as assisting in the reduction of financial costs and the amount of time spent on maintenance. Predictive maintenance has largely become available due to the increased interest in industrial intelligence alongside technological innovation, leading to an expansion of data collection and rapidly evolving data sources.

The predictive maintenance approach consists of utilizing data that can diagnose a system's current condition and analyze future maintenance needs (Pech et al., 2021). Machine learning (ML) and artificial intelligence (AI) can use current and historical data on the state of a vehicle and peer vehicles in its fleet to power this maintenance approach. This is achieved through AI and ML's statistical methodologies that capture underlying patterns in sensor and maintenance data that inform predictions on a vehicle's current and future state. For instance, data from disparate sensor sources can be fed into unsupervised ML models which can identify anomalies in a vehicle's current state compared to historical data, which can in turn identify a maintenance need (Kellner et al., 2021). Additionally, when known faults in the historical dataset are labeled, supervised ML methods such as classification can be trained to identify and classify future system faults based on a vehicle's current sensor data.

Pairing predictive maintenance along with an intuitive dashboard can help automate maintenance by making it more efficient and giving stakeholders the necessary information to make crucial decisions to maintain military readiness. Predictive maintenance solutions that include decision support reduce the amount of cognitive load on users and provide point-of-need information and guidance that allow for a better-informed decision-making process. Maintenance solutions which include low system restrictiveness allow the users to better control how they make their decisions (Kellner et al., 2021). Such intuitive dashboards are necessary for an organization adopting a predictive maintenance approach because these maintenance decisions are ultimately the responsibility of the leaders of the organization. Ensuring the data is shown in a variety of ways will assist with decision-making, reduce cognitive load on users, and keeps the human in the loop with AI.

Predictive maintenance is an innovative solution for organizations needing to maintain readiness due to its ability to coalesce data from disparate sources including routinely collected maintenance data and onboard systems data. The large amount and varied nature of these data sources necessitates a shift in focus during record keeping when it comes to equipment maintenance. For example, shifting from paper receipts and records to a digital database for vehicle maintenance activities, which is exhibited by many industry mechanic shops today. This shift toward digital records allows a historical view into work that was previously done on a vehicle. Additionally, more automatic, and real-time approaches to data collection have become available through the invention and improvement of communication networks. For example, the Controller Area Network (CAN) serial communications bus, referred to as "CAN bus", allows communications for in-vehicle networks often utilizing a collection of microcontrollers. In the 1990s, major vehicle manufactures started to adopt the CAN bus technology which has experienced significant growth leading to wide adoption (Davis et al., 2007).

This paper outlines the benefits, considerations, and tradeoffs of different algorithms and user-centered design approaches in predictive maintenance vehicle. Specifically, this paper considers three AI/ML approaches to predictive maintenance: survival regression, anomaly detection, and classification. Finally, a case study is provided is provided of these predictive maintenance solutions applied to one vehicle maintenance problem and dataset, along with the creation of a decision support dashboard to enable its usage in the field.

REVIEW OF PREDICTIVE MAINTENANCE MODELING APPROACHES

As previously noted, predictive maintenance has been widely used across industries to improve upon standard maintenance scheduling and reduce time, expense, and downtime due to component failure. These approaches continue to grow in their usefulness as more powerful ML techniques and greater amounts of data are collected and able to be easily stored (Kaur et al., 2018; Selcuk, 2017; Susto et al., 2015; Theissler et al., 2021; S. Wu, 2011). The DoD has especially prioritized a shift to predictive maintenance solutions as emerging technologies have enabled such strategies to make U.S. forces more efficient, effective, and superior to potential adversaries (Bell, John, 2008). In this section, a review is provided of the data processing and predictive modeling approaches in the scientific literature used to generate robust and reliable predictive maintenance solutions with a specific focus on vehicles.

Data Preprocessing and Feature Building

Predicting upcoming faults or repairs from sensor data first requires careful feature engineering through pre-processing of sensor data. In some cases, features can essentially be used in their raw format. Engine coolant temperature, for instance, can provide useful information on faults that may occur when an engine is consistently heated beyond its normal operating temperature range, as well as faults that could cause excessive coolant temperature in the first place. As such, raw values of coolant temperature can be a highly predictive feature on its own with minimal processing needed. For other features, however, it can be beneficial to threshold, normalize, or combine features in linear or nonlinear ways to construct new features that may be more indicative of faults than their raw values. For instance, comparing the difference or delay between a vehicle's drive-by-wire (DBW) commands and the output could indicate correct or malfunctioning system and mechanical responses.

One common transformation is to analyze features in the frequency domain – that is, viewing how powerfully they are oscillating at particular frequencies. For example, when vehicle components are resonating at a particular frequency, this can be indicative of an underlying failure, or can cause damage to the component itself, or neighboring components if the resonance continues unabated. Such analyses are typically conducted through Fourier or wavelet transforms on the data, a process that can give the power at frequencies of interest over time, and has been successfully used in many predictive maintenance applications (Bonnevay et al., 2020; S. S. Patil & Gaikwad, 2013; Sáinz-Pardo Díaz, 2021; Zabihi-Hesari et al., 2019).

For all features, even those that do not need to be transformed, there are other pre-processing considerations that can be used to reduce noise and improve predictive power. For instance, missing data is very common from physical sensors (Rafsunjani et al., n.d.). Different modeling approaches have different levels of robustness to missing data, and various data imputation methods can be used to fill in these missing values. These methods range from simply replacing all missing values with the sensor's mean value or zero, to more complex approaches such as linearly or nonlinearly interpolating missing data from neighboring points, or building separate statistical models to impute missing data from other features (Luengo et al., 2012). Depending on what has caused data loss to occur (whether it is missing at random or is correlated with other states of the system, such as faults), the approach to handling missing data can drastically affect modeling outcomes and must be handled carefully (Bhaskaran & Smeeth, 2014). Another important preprocessing step can be to smooth or down-sample features before inputting them into a statistical model. High frequency time series can be subject to high levels of noise and may incur high computational costs due to data size. In such cases, reducing the sampling frequency of the data can shorten computation time, decrease storage needs, and maintain or even increase a model's predictive ability.

Modeling Approaches

Once useful features have been pulled from the data, a statistical model can be built to determine underlying patterns that predict faults or repairs before they occur. A wide variety of statistical and ML modeling approaches have been successfully utilized in vehicle and other manufacturing predictive maintenance solutions (Seera et al., 2014; Theissler et al., 2021). Through a survey of the literature, three historically successful classes of analysis have been identified: anomaly detection, classification, and survival regression.

Anomaly Detection

Anomaly detection is the simplest of these three major classes, whereby historical and peer vehicle data is used to identify joint distributions of operational sensor data that are typically observed during normal working condition of the vehicle. If a set of observations falls significantly outside of this distribution, this constitutes an anomaly in the which may indicate a fault or early warning sign of an incoming fault. Such methods have proven highly successful in identifying faulty and atypical modes of operations in vehicles (Khan et al., 2019; Theissler et al., 2021).

Anomaly detection is typically achieved through unsupervised ML algorithms, and such approaches have some major advantages over the supervised methods discussed in upcoming sections. Namely, since they do not require labels on anomalous versus normal conditions or specific fault types, the data needed for them is typically more easily acquired (e.g., for vehicle predictive maintenance, simply through interfacing with a vehicle's CAN bus with no manual labeling required). This also means that these methods are generalizable to any sort of anomaly – they need not be trained on a particular fault but can simply detect any major deviation from normal working condition. As such, this approach is more likely to detect very rare faults than a supervised model, because if a fault never or rarely occurs in a dataset, it is impossible to train a supervised model to detect it. With unsupervised anomaly detection, if that fault is reflected by significantly deviant sensor data, an anomaly detection model with sufficient training data can identify it.

As a tradeoff for their ease in data collection and training, anomaly detection models provide the least specific predictions on what fault lies behind the anomalous data, as these data are not known to the model. In its simplest form, this can lead to the model simply reporting that some fault has (or will soon) likely occur without reporting what that fault is, which may not be sufficient information for a maintainer to act upon. One of the most straightforward approaches to anomaly detection is the Mahalanobis Distance (MD). MD is very similar to the intuitive and well-known Euclidian distance, but the MD takes into consideration the linear correlations between variables in a high-dimensional feature space (De Maesschalck et al., 2000). Using the MD, one can determine a score for a data point indicating how far it is from the typical distribution (Hu et al., 2016). In a predictive maintenance setting, this means that each point in this space represents the complete set of sensor features over a short period of time, and these data can come from historical readings from the vehicle of interest as well as peer vehicle data. Then, during vehicle operation, the vehicle's current state can be projected into this space, and the MD calculated to determine its distance from normal operation. The MD provides a score which can be thought of as a generalization of the number of standard deviations this data point falls from the historical distribution. If this value passes a threshold, the current state is classified as an anomaly. This technique has been widely used to identify failure modes in manufacturing and electronics (N. Patil et al., 2015; Wang et al., 2013) and for vehicle predictive maintenance from on-board sensors (Lin et al., 2010) where MD was used to detect anomalous sensor data in unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs). Wen and Guo use a weighted version of the MD to produce estimates on failure times of ball-screw systems used widely in machinery (Wen & Gao, 2018).

An alternative nonlinear method for anomaly detection is the one-class support vector machine (OCSVM), a nonparametric boundary-based method for identifying outliers in a dataset. It is derived from the classical support vector machine (SVM), a classification method by which a boundary is drawn around all points of each class in the high-dimensional feature space, allowing future points to be classified by which boundary they are within or closest to. An OCSVM does this for a single class of points, allowing it to be an unsupervised method with no class labels. This method enables anomaly detection when this boundary is allowed to leave some extreme points outside of the class – such points, and all future points falling outside this boundary, are labeled as outliers.

OCSVM models are simple to compute as they do not consider most of the data that sits near the center of the high-dimensional distribution. This, however, also makes them very sensitive to extreme points, and extensive tuning of hyperparameters and cleaning of data can be necessary to enable their effective usage. While less popular, OCSVM models have also seen success in predictive maintenance. Most notably, Guo et al. trained a modified OCSVM which better dealt with outliers in the training data through a density-based optimization algorithm and trained this model on UAV sensor data. Their model effectively identified manually inserted faults in a high fidelity simulation of UAV operation (Guo et al., 2019). However, SVM-type models are typically outpaced by other models when large amounts of data are available.

Survival Regression

Analyzing data through time series regression falls on the opposite end of the spectrum of model complexity and precision compared to anomaly detection. By continuously analyzing sensor data as an ordered series of time points, one can achieve much more fine-tuned predictions such as the RUL of particular components. Estimating such events can be challenging, requiring large amounts of clean data and careful cross-validation to ensure adequate model performance. When effective, however, time series methods can provide the most specific and predictions and recommendations for predictive maintenance by providing a precise estimate of RUL (Hivarekar et al., 2020). Such predictions, if able to be utilized, can enable the most optimal repair schedule so that downtime can be reduced; components can be used to as near their maximum lifetime as feasible, and supply chain logistics can be best optimized by ensuring necessary parts are acquired in time for the required maintenance without unnecessary storage time or transport to new locations.

Survival regression is a statistical modelling technique for time series that predicts the expected timing of future events. This approach has been widely used in medicine, where it can be used to predict the probability of patient survival or adverse medical events within an upcoming time window. These techniques have since been widely used in manufacturing to predict component failures and estimate RULs. A key feature of survival methods is their correction for “censoring” in data. Censored events are those which occurred before or after the data collection period for that individual. For example, if a vehicle stops appearing in a dataset, it may be because the vehicle had been replaced by a newer model, or it may be that the vehicle encountered a catastrophic failure that forced its retirement, or it may be that the vehicle is still in use and new data simply isn’t available yet. These unknowns about when and whether adverse events appear or have been censored in our data can lead to significant biases in predictions if they are not accounted for. Survival methods are specifically designed to correct for such biases.

Survival analysis spans a wide range of methods, but for regression, the most common model is the cox proportional hazard model. This model predicts a baseline hazard – the likelihood of failure – for a component, and then proportionally modifies the hazard based on the sensor data from that vehicle. A less common alternative is the accelerated failure time (AFT) model, which learns how sensor data may accelerate or decelerate the failure of a component. These models serve similar purposes but make different assumptions on the mathematics linking the sensor covariates with component failure events. In relevant literature, de Almeida Costa et al. (2021) used a Cox proportional hazard model to derive survival curves of train wheel wear from 17 years of historical data and identified optimal repair strategies to minimize downtime and failure (de Almeida Costa et al., 2021). AFT models have not been used as widely for vehicle predictive maintenance, but Syamsundar, Naikan and Couallier utilized an AFT model for predictive maintenance of a cement plant, using sensor data to effectively predict the RUL of cement roller mills and identify the optimal maintenance strategy for them (Syamsundar et al., 2020).

Classification

Classification methods are the most well-studied and successful approaches in modern machine learning. Given a set of data points labeled in two or more discrete classes, these methods learn the underlying features best differentiating the classes and allow novel, unlabeled data to be sorted into these classes or given scores for their similarity to each class. For predictive maintenance, classification methods look at sensor data around the times when there is no fault (the “Normal” class) and sensor data just before and during the time when a fault has been recorded (the “Fault” class) (Theissler et al., 2021). Often, a more granular understanding of the fault is desired, and additional classes can be created for specific types of faults. This effort considered two main types of classification models: logistic regression and tree-based models.

Logistic regression is a form of generalized linear regression whereby a logistic curve is fitted to binary data. For prediction, a novel data point is assigned a probability between 0 and 1, and a chosen threshold is used to determine whether that point belongs in the “normal” (0) or “faulty” (1) class. Logistic regression is one of the simplest forms of classification, making assumptions on the linear relationship between features and the presence or absence of faults, but the coefficients the model fits are highly interpretable, allowing for a clearer understanding of which features are indicative of which types of faults (Phillips et al., 2015).

Decision tree-based models are some of the most popular and flexible class of nonlinear classification models. A decision tree is a series of binary decisions, each of which is a threshold which splits the data into two groups which most highly separate the training data by its class labels. This recursive process continues until the data is completely separated into individual classes or predetermined stopping points are reached. The result is essentially a flowchart by which unlabeled data can easily be classified – the prediction algorithm simply determines whether a new data point’s features are greater than or less than each threshold until it reaches a point in the tree at which all the data comprises a single class (or is dominated by a single class) and predicts that the novel data also belongs to that class.

Because of this simple formulation, decision tree models are trained very rapidly. They also are highly explainable – if one wishes to know why a data point was sorted into a particular class, they can easily see which features led to this decision. However, this simple process also has several drawbacks. For one, it can be overfit to training data and does not necessarily generalize well, especially when certain classes are especially rare. The training algorithm is also “greedy” – at each step, it chooses the best threshold to split on, even if a series of individually less optimal splits might lead to a more optimal final classification. To combat this, ensembles of trees termed “forests” are typically used instead, whereby many decision trees are trained each on subsets of the data and features. From there, each tree provides one vote for the classification of a novel data point. This helps prevent the algorithm from overfitting and getting stuck in local minima due to the greedy nature of the basic algorithm.

Prytz et al. use the simplest of these ensemble methods, the random forests classifier, to predict if an air compressor replacement is needed at the current inspection time or if it is sufficiently likely to survive until the next inspection. In their use case of commercial truck and bus maintenance, such repairs could only occur at these regular maintenance intervals, making this classification method a superior solution to a more fine-tuned remaining useful life (RUL) prediction in this situation (Prytz et al., 2015). Cerqueira et al. use an even more sophisticated modification on decision trees termed gradient boosted trees (GBTs) where a forest of trees is built by constructing one decision tree which weakly fits the data, then constructing a second tree which fits to the residuals of that tree (i.e., the signals that the first tree was unable to learn), followed by more trees successively fitting to the remaining residuals. This is a slower and more computationally taxing process but yields a robust and generalizable statistical model. These authors used GBTs trained on carefully engineered features to predict whether a truck failure was due to a fault in its air pressure system or another unrelated component (Cerqueira et al., 2016).

CASE STUDY

Determining which modeling approach to apply for predictive maintenance is largely dependent on the goal of the application and the data available. The use case selected for this effort focuses on the U.S. Army vehicle maintenance sector, specifically for Battalion Maintenance Officers (BMO). Currently, the BMO process involves scheduled time-based maintenance determined by technical specifications for each component. Additionally, maintainers submit faults that are reviewed by the BMO, however after initial submission and review, there is little visibility into the status of those faults. BMOs only receive readiness of the fleet data as it is reported monthly from the lowest field level maintenance asset for an operational view.

A sample set of data was provided by the U.S. Army Combat Capabilities Development Command (DEVCOM) and obfuscated for non-military usage by obscuring vehicle models, VINs, and component serial numbers to remove information about specific vehicles and components. These data consist of two main datasets: 1) Onboard vehicle sensor data, and 2) Maintenance data. The onboard vehicle sensor data includes time series recordings from the vehicles' CAN bus. These include 1 Hz sensor data, fault code event timestamps, and startup data. The Maintenance dataset provided a history of faults recorded by maintainers, including descriptions of faults that occurred, if and how they were addressed, and whether the inspection was scheduled or not.

Both the maintenance and sensor datasets were collected during a two-and-a-half-year timespan. It is important to note that much of the data does not span this entire time period, and many vehicles have left and/or right censored data. As such, it required care to ensure biases were not inserted into the data due to this censoring. If RUL estimates are computed using these data, they will need to take censoring biases into account, which requires careful distinction between what is censored vs non-censored observations.

When implementing data modeling for predictive maintenance, there are a few considerations to keep in mind: most importantly, the types of data, their completeness, and how the goal of the data model can be framed to utilize the available data. Given the limited number of vehicles and narrow period in our dataset, we identified the amount of non-censored data available for predicting component failure. The search was restricted to only those parts costing over $100 to avoid small, non-critical components such as individual screws, and to maximize the cost savings that could be applied by ensuring that these expensive components are used until close to the end of their useful life before replacement. Only part replacements, rather than repairs, were considered, as full replacements of components were significantly more common than repairs. Finally, forward predictions for any replacements done during annual inspections were excluded (approximately 25% of replacements), as it was impossible to discern from this dataset whether a component was replaced due to its condition, or if it was replaced preemptively in a schedule-based maintenance approach (which would bias the RUL estimates if such unnecessary replacements were included in the training set).

This review of the data yielded four components with at least 70 replacements across the entire dataset, outlined in **Table 1**. Additionally, three different, but related caterpillar engines were combined, which brought their total replacements above 70 as well.

Table 1. Summary data on most frequently replaced components in database.

Component	# Replaced	In # Vehicles	In # Models	% Non-Censored	Cost (USD)
Diesel Engine*	118	106	11	10%	**$52,994
Electronic Housing	78	73	6	6%	$9,212
Voltage Regulator	73	67	12	8%	$426
Alternating Generator	70	62	6	11%	$3,159
Combat Turret	70	63	6	10%	$35,059

**The three most frequently used types of diesel engines were combined for this component*
***Due to the combination of three engine models, the replacement cost is variable, the lowest being $39k*

MODEL SELECTION

Based on the available data and use case, the previously described modeling approaches were evaluated. Each modeling approach and its relevance to the selected use case are described below, including pros, cons, and ultimate reasons for selection or rejection.

Survival Regression

Survival regression is a powerful approach due to its ability to generate to RUL estimates and the more precise insight RUL can provide for predictive maintenance. However, while survival methods are specifically designed to control for the statistical biases introduced by censored data, predictions can be highly variable and imprecise in datasets with a large proportion of censored data, and it is recommended that data not have greater than 50% censored data when using survival methods. Additionally, survival methods require relatively continuous data leading up to failure events to build a model which can reliably predict the time to event failure in the future. Most vehicles in the available subset had sparse data, which would lead to these highly variable and inconsistent RUL predictions. Thus, survival regression was not considered further for this case study.

Anomaly Detection

In exploring the potential use for anomaly detection with the dataset available, unsupervised anomaly detection became relevant and applicable for the CAN bus sensor data given the unknown nature of when exactly the cause of the fault occurred. Unsupervised anomaly detection was used to identify extraordinary and potentially faulty sensor data during operation. Some initial testing was done using the unsupervised approach of Isolation Forests but showed little predictive capability when applied to overall days of operation. A challenge when dealing with a vehicle is that there are states of idleness and movement that can cause expected spikes in a sensor's operation. For example, machines connected to an Internet of Things (IOT) in a manufacturing plant regularly work at a consistent rate for a 24/7 period, providing a robust dataset for AI/ML algorithms. These more consistent rate sensors would do a better job using unsupervised anomaly detection over a start/stop type of machine.

An alternative to anomaly detection commonly implemented is more traditional change detection techniques which range in methods such as onset detection for moments of elevation, peak detection above a defined criteria, and change detection through monitoring the trend and expected drift. To work within a more defined set of parameters, various change detection techniques were implemented for featurization; with plans to revisit unsupervised anomaly detection for finding potential outliers in more defined periods. While identifying faulty sensor data for a vehicle provides insight for the moment and may indicate an impending failure that can be acted upon, it does not paint a clear picture of when a failure is going to occur. For this case study, where we sought a more precise indication of the timeline for failure, we opted to focus on a supervised model as the main driver, with the potential to add unsupervised anomaly detection as an additional module later to identify rarer events which were not present in large numbers in the available datasets.

Classification

While a classification approach cannot predict a precise RUL, it can be used to predict whether a component is likely to fail in a pre-determined length of time (e.g., one week). As such, the issues regarding censored data are much less pronounced when using classification methods. Unlike with survival regression, it is not important how long it has been since a component was last replaced and how long it will take before it fails after the cessation of data collection. This increases the utility of more of the dataset because data can reasonably be given normal labels without introducing statistical bias into the model.

However, the low number of replacements in this dataset makes this a highly unbalanced classification problem. In operation, a vehicle will be in normal working condition far more often than it will have faults, and therefore, data labeled as faulty are far outnumbered by data labeled as normal. Fortunately, there are strategies in classification methods to account for this imbalance. Using correct metrics for the problem (such as precision and recall rather than a simple accuracy measurement) ensures that the best model does not always predict that a vehicle's operation is normal simply because it is the most common class. These considerations can also be built into the model's loss function to ensure it learns the difference between these highly unbalanced classes. Additionally, classification models do not necessarily need to present a binary normal or faulty classification; instead, models like logistic regression compute a probability that a period of onboard vehicle data should fall into one class or the other. This probability can be utilized in human decision-making and can be compared against the base rate to determine a more meaningful risk of failure.

For example, consider a situation where only 0.1% of sensor data in the training set is labeled faulty, and a logistic regression model returns a 2% chance that the current sensor data are predicting a fault within the next seven days. There are two competing factors to weigh. On one hand, at a 2% failure chance, it is still unlikely that a fault will occur, simply because faults are always a rare event. However, this 2% indicates a twenty-fold risk increase over the typical chance of a fault occurring in the next seven days. Depending on a mission's timeline, the current needs for the vehicle, and potential solutions (e.g., immediate replacement, replacement at the next destination, switching out with a backup vehicle, etc.), these probabilities, rather than rote binary classification, can provide actionable information that respects both the inherent statistical imbalances in predictive maintenance and the contextual factors surrounding a maintenance decision. Table 2 provides a summary of the pros and cons for each modeling approach along with reasons for selection or rejection.

Table 2. Pros and cons of ML models for predictive maintenance and selection reasoning for case study

Model Type	Pros	Cons	Selection Reasoning
Survival Regression	• RUL estimation • Censorship handling • Formulated specifically for event detection in time series data	• Requires a high ratio of non-censored data. • Larger periods of time may be required if the failure event is infrequent. • Assumptions must be made on relationship between features and hazard function	• Not selected, due to current dataset constraints • Limited non-censored observations within dataset.
Anomaly Detection	• Unsupervised & semi-supervised options require no or minimal data labels • Detects any type of fault that is sufficiently different from normal operation	• Large number of ‘normal’ observations required to correctly separate out outliers. • No indication of type of fault detected. • No indication of timeline for impending fault to occur.	• Selected for future addition to complement primary classification model to identify rare outlier faults. • Using change detection as a feature engineering method for the chosen classification approach
Classification	• Many methods available, such as logistic regression, or decision-tree based methods • Can provide prediction probability • Can define the time period of interest over which an impending fault is identified, which can be variable across components	• Time is not considered with the level of granularity as survival or other time series methods • RUL estimation is not possible – only probability of failure within set time period	• Selected • Allows for greater flexibility in defining problem when met with highly censored data.

Time series variance was first detected using simple statistical transformations common in change detection, including sequential analysis, peak detection, and onset detections, to indicate when individual points in the CAN bus time series data fell outside their normal range. Then, initial classification models were built using a logistic regression classifier implemented in the SciKit-Learn library in Python and used the features described in the previous onboard vehicle data section as regressors, along with features derived from the change detection approaches. Logistic regression is a generalized linear regression approach that provides coefficients and p-values for each input feature regarding how well and in what direction they inform a classification of sensor data as normal or faulty. This allows for a more easily interpretable model, which is beneficial in the initial feature engineering and modeling process. At a later stage in development, once highly predictive features and a basic pre-processing and modeling strategy have been solidified, investigation will center on using a more complicated nonlinear model, such as a tree-based method, which typically gives more accurate predictions at the expense of some interpretability. The SciKit-Learn package makes transitioning between these model types simple through its object-oriented approach to statistical models.

DECISION SUPPORT DASHBOARD

In order to visualize the predictions made by the AI/ML models along with other useful data from the sample dataset, a decision support dashboard was designed and developed. This web- or tablet-based tools allows BMOs and other decision makers to view the readiness of their battalion at-a-glance.

Multiple user interviews were conducted with US Army maintenance personnel to learn more about the needs of the end-users which informed the dashboard design. Data deemed necessary for inclusion in the dashboard were components with a high probability of failure, faults with a high occurrence probability, actual faults versus predicted faults, items that require maintenance, failure frequency and average downtime. Several critical parameters selected for optimization were time, money, manpower, parts on hand, the number of vehicles available for battle, and the ability to visualize actionable information that soldiers can work with to improve these parameters.

To meet user needs, the dashboard was designed to report implications of part prognosis (e.g., if part *W* fails, component *X* will be implicated, and System *Y* will be down for an estimated *Z* number of days). This assists in user decision-making with respect to what vehicles are available for use while others undergo maintenance. The dashboard also includes mission readiness and effectiveness reports (e.g., *X* vehicles are ready to be deployed, *Y* are undergoing maintenance, *Z* are down and awaiting parts, etc.). These reports allow the BMO to see an overview of battalion readiness. Another component of the decision support dashboard is recommendations of part procurement (e.g., part *X* is predicted to fail within the next 30 days so it may be important to procure a spare). This assists users in making decisions of what parts are necessary to order. Lastly, there are features that assist in the refinement of the Maintenance Process Cycle (e.g., In quarter 4, *X* scheduled maintenance procedures were performed which found no presenting issues; this maintenance cycle may need to be revalued for efficacy). Figure 1 shows the initial design of the decision support dashboard.

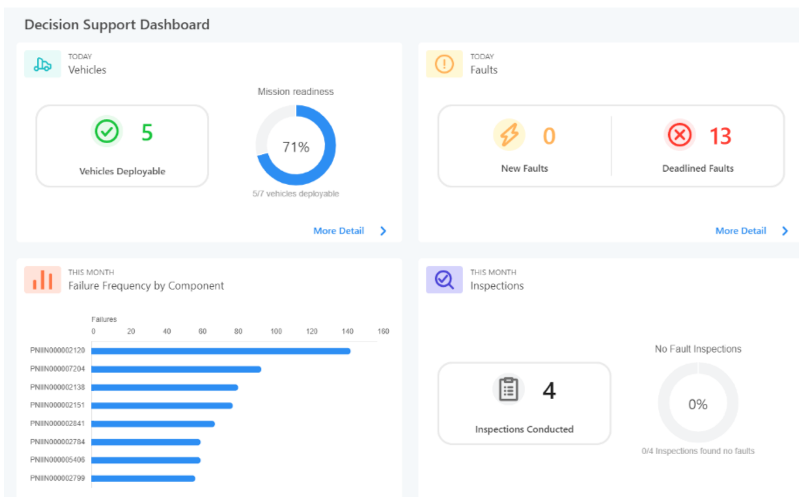

Figure 1. Decision Support Dashboard Design

RESULTS

Following the development of the initial decision support dashboard, preliminary evaluations were conducted with three representative end-users (2 male, 1 female). One of the end-users was a retired Air Force maintenance officer, while the other two had background in maintenance, but to a lesser degree. Evaluations were conducted using a series of four tasks requiring the use of the initial dashboard.. Throughout the evaluation, users were asked to speak out loud about what they were thinking and feeling about each step required to complete the tasks. Following the tasks, users were probed with additional questions about their thoughts on the system and completed the System Usability Questionnaire (SUS), which is a way to measure perceived usability, and the Net Promoter Score (NPS) questionnaire, is a single survey question that rates the likelihood that a user would recommend a product to a friend.

Overall, the preliminary system design feedback was positive. Users commented that the navigation was very easy to understand, the breadcrumbs were useful, and they enjoyed the clean interface look. On average the SUS score was 70.8 which puts the system in the Above Average category (< 68 is considered Above Average). The NPS Score was 60, indicating an excellent system. Areas of improvements were also identified including, adding a snapshot view of the three components most likely to fail in the near term, adding descriptors to orient the user, and replacing some charts with less complex versions.

DISCUSSION

Implementation of predictive and prognostic maintenance approaches require careful preparation and handling of maintenance data. The most important consideration when determining what modeling approach to take for predictive and prognostic maintenance often depends on the historical data available. For very complete and robust datasets, RUL estimation is the ideal prediction target made available by survival regression. However, in order for RUL predictions to be accurate, the dataset needs to contain many instances of observed event lifespans, including time of installation, time to the first use, and time until failure. When RUL estimation is not achievable due to limited historical failures or incomplete datasets, then anomaly detection and classification approaches are more suitable since they can be used to predict the likely failure time of a particular component. The biggest challenge in applying a classification approach is determining which events should be classified for prediction. Vehicle maintenance may be completed for a variety of reasons, and it can be difficult to discern if a component was replaced due to failure or as part of a time-based approach when the data is unclear. The rarity of event failure in relation to what is viewed as an observation for prediction can create challenges in a largely imbalanced training and testing set. The case study in the current effort provided the opportunity to conduct a research process and gain useful information to narrow down the most effective AI/ML algorithms and provide insight into the ways these algorithms can be applied for other use cases.

Another challenge facing the general use of predictive and prognostic maintenance models is the lack of trust by users. One of the most effective ways of increasing model accuracy and trust, is keeping the human in the loop. Including human input (such as maintainer entered vehicle faults) to the ML modeling approach increases the efficacy of the model and provides the users a sense of trust, as well as providing a more direct tracking of failures. Allowing the human to remain in the loop also allows training of an accurate predictive maintenance model with a minimum cost for the organization by using the user's experience and knowledge (X. Wu et al., 2022). Preliminary evaluations of the decision support dashboard developed for this effort indicate the system was above average for usability. Additional iterative user testing will be undertaken as the effort continues to ensure the human remains in the loop and to enhance the generalizability of the results.

CONCLUSIONS

This paper evaluated several ML approaches for building a predictive maintenance tool, using a dataset provided by US Army DEVCOM, for the purpose of increasing the flow of information and the efficiency of their BMO processes. Of the approaches evaluated, survival regression was ruled out due to the limited lifespan and small number of failures in the sample dataset provided. Anomaly detection was evaluated to build features for further processing in a classification model but was ultimately discarded. However, future efforts could consider anomaly detection with larger, more robust datasets. Classification, by logistic regression, was applied in this effort for predicting the likely failure of a component within a time window by understanding the sensor characteristics surrounding a component failure. Many challenges arise when working with real-world datasets, not least of which is that real-world vehicle maintenance data may be censored or lacking in completeness. Reasons for incomplete data may vary, for example, if data collection begins after a vehicle has been in use, or maintenance records may not be completed appropriately when components are replaced. It is important to note however, that even though these data, which are important for the proper training on a ML algorithm, are not always available, there are still ways to incorporate AI/ML into predictive maintenance. There are additional challenges in dealing with sensor data of vehicles, such as wider variance in starting and stopping states which may not be seen in other larger datasets. Plans for the current effort include refinement of the selected algorithms, further user testing with US Army maintenance personnel to gain more insight into the validity of the chosen model and the usefulness of the decision support dashboard in day-to-day tasks, and full development of the decision support dashboard.

REFERENCES

Bell, John. 2008. "Condition Based Maintenance Plus DoD Guidebook."

Bhaskaran, Krishnan, and Liam Smeeth. 2014. "What Is the Difference between Missing Completely at Random and Missing at Random?" *International Journal of Epidemiology* 43(4):1336–39. doi: 10.1093/ije/dyu080.

Bonnevay, Stéphane, Jairo Cugliari, and Victoria Granger. 2020. "Predictive Maintenance from Event Logs Using Wavelet-Based Features: An Industrial Application." Pp. 132–41 in *14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019)*, *Advances in Intelligent Systems and Computing*, edited by F. Martínez Álvarez, A. Troncoso Lora, J. A. Sáez Muñoz, H. Quintián, and E. Corchado. Cham: Springer International Publishing.

Davis, Robert I., Alan Burns, Reinder J. Bril, and Johan J. Lukkien. 2007. "Controller Area Network (CAN) Schedulability Analysis: Refuted, Revisited and Revised." *Real-Time Systems* 35(3):239–72. doi: 10.1007/s11241-007-9012-7.

De Maesschalck, R., D. Jouan-Rimbaud, and D. L. Massart. 2000. "The Mahalanobis Distance." *Chemometrics and Intelligent Laboratory Systems* 50(1):1–18. doi: 10.1016/S0169-7439(99)00047-7.

Guo, Kai, Liansheng Liu, Shuhui Shi, Datong Liu, and Xiyuan Peng. 2019. "UAV Sensor Fault Detection Using a Classifier without Negative Samples: A Local Density Regulated Optimization Algorithm." *Sensors* 19(4):771. doi: 10.3390/s19040771.

Hivarekar, Nita, Swarup Jadav, Vivek Kuppusamy, Priyanka Singh, and Chanchal Gupta. 2020. "Preventive and Predictive Maintenance Modeling." Pp. 1–6 in *2020 Annual Reliability and Maintainability Symposium (RAMS)*.

Hu, Zewen, Mingqing Xiao, Lei Zhang, Shuai Liu, and Yawei Ge. 2016. "Mahalanobis Distance Based Approach for Anomaly Detection of Analog Filters Using Frequency Features and Parzen Window Density Estimation." *Journal of Electronic Testing* 32(6):681–93. doi: 10.1007/s10836-016-5623-z.

Kaur, Karamjit, Matt Selway, Georg Grossmann, Markus Stumptner, and Alan Johnston. 2018. "Towards an Open-Standards Based Framework for Achieving Condition-Based Predictive Maintenance." Pp. 1–8 in *Proceedings of the 8th International Conference on the Internet of Things*, *IOT '18*. New York, NY, USA: Association for Computing Machinery.

Kellner, Domenic, Maximilian Lowin, Moritz von Zahn, and Johannes Chen. 2021. *Towards Designing a User-Centric Decision Support System for Predictive Maintenance in SMEs*. Gesellschaft für Informatik, Bonn.

Khan, Samir, Chun Fui Liew, Takehisa Yairi, and Richard McWilliam. 2019. "Unsupervised Anomaly Detection in Unmanned Aerial Vehicles." *Applied Soft Computing* 83:105650. doi: 10.1016/j.asoc.2019.105650.

Lin, Raz, Eliyahu Khalastchi, and Gal A. Kaminka. 2010. "Detecting Anomalies in Unmanned Vehicles Using the Mahalanobis Distance." Pp. 3038–44 in *2010 IEEE International Conference on Robotics and Automation*.

Luengo, Julián, Salvador García, and Francisco Herrera. 2012. "On the Choice of the Best Imputation Methods for Missing Values Considering Three Groups of Classification Methods." *Knowledge and Information Systems* 32(1):77–108. doi: 10.1007/s10115-011-0424-2.

Patil, Nishad, Diganta Das, and Michael Pecht. 2015. "Anomaly Detection for IGBTs Using Mahalanobis Distance." *Microelectronics Reliability* 55(7):1054–59. doi: 10.1016/j.microrel.2015.04.001.

Patil, S. S., and J. A. Gaikwad. 2013. "Vibration Analysis of Electrical Rotating Machines Using FFT: A Method of Predictive Maintenance." Pp. 1–6 in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*.

Pech, Martin, Jaroslav Vrchota, and Jiří Bednář. 2021. "Predictive Maintenance and Intelligent Sensors in Smart Factory: Review." *Sensors* 21(4):1470. doi: 10.3390/s21041470.

Phillips, J., E. Cripps, John W. Lau, and M. R. Hodkiewicz. 2015. "Classifying Machinery Condition Using Oil Samples and Binary Logistic Regression." *Mechanical Systems and Signal Processing* 60–61:316–25. doi: 10.1016/j.ymssp.2014.12.020.

Rafsunjani, Siam, Rifat Sultana Safa, Abdullah Al Imran, Shamsur Rahim, and Dip Nandi. n.d. "International Journal of Information Technology and Computer Science(IJITCS)." *International Journal of Information Technology and Computer Science(IJITCS)* 11(2):21.

Sáinz-Pardo Díaz, Judith. 2021. "Predictive Maintenance and Spectral Analysis: From Fourier to Machine Learning."

Seera, Manjeevan, Chee Peng Lim, Saeid Nahavandi, and Chu Kiong Loo. 2014. “Condition Monitoring of Induction Motors: A Review and an Application of an Ensemble of Hybrid Intelligent Models.” *Expert Systems with Applications* 41(10):4891–4903. doi: 10.1016/j.eswa.2014.02.028.

Selcuk, Sule. 2017. “Predictive Maintenance, Its Implementation and Latest Trends.” *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 231(9):1670–79. doi: 10.1177/0954405415601640.

Susto, Gian Antonio, Andrea Schirru, Simone Pampuri, Seán McLoone, and Alessandro Beghi. 2015. “Machine Learning for Predictive Maintenance: A Multiple Classifier Approach.” *IEEE Transactions on Industrial Informatics* 11(3):812–20. doi: 10.1109/TII.2014.2349359.

Syamsundar, A., V. N. A. Naikan, and V. Couallier. 2020. “Accelerated Failure Time Models with Corrective and Preventive Maintenance for Repairable Systems Subject to Imperfect Repair.” Pp. 143–52 in *Reliability, Safety and Hazard Assessment for Risk-Based Technologies, Lecture Notes in Mechanical Engineering*, edited by P. V. Varde, R. V. Prakash, and G. Vinod. Singapore: Springer.

Theissler, Andreas, Judith Pérez-Velázquez, Marcel Kettelgerdes, and Gordon Elger. 2021. “Predictive Maintenance Enabled by Machine Learning: Use Cases and Challenges in the Automotive Industry.” *Reliability Engineering & System Safety* 215:107864. doi: 10.1016/j.ress.2021.107864.

Verhagen, Wim J. C., and Lennaert W. M. De Boer. 2018. “Predictive Maintenance for Aircraft Components Using Proportional Hazard Models.” *Journal of Industrial Information Integration* 12:23–30. doi: 10.1016/j.jii.2018.04.004.

Wang, Yu, Qiang Miao, Eden W. M. Ma, Kwok-Leung Tsui, and Michael G. Pecht. 2013. “Online Anomaly Detection for Hard Disk Drives Based on Mahalanobis Distance.” *IEEE Transactions on Reliability* 62(1):136–45. doi: 10.1109/TR.2013.2241204.

Wen, Juan, and Hongli Gao. 2018. “Remaining Useful Life Prediction of the Ball Screw System Based on Weighted Mahalanobis Distance and an Exponential Model.” *Journal of Vibroengineering* 20(4):1691–1707. doi: 10.21595/jve.2018.19099.

Wu, Shaomin. 2011. “Preventive Maintenance Models: A Review.” Pp. 129–40 in *Replacement Models with Minimal Repair, Springer Series in Reliability Engineering*, edited by L. Tadj, M.-S. Ouali, S. Yacout, and D. Ait-Kadi. London: Springer.

Wu, Xingjiao, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. “A Survey of Human-in-the-Loop for Machine Learning.” *Future Generation Computer Systems* 135:364–81. doi: 10.1016/j.future.2022.05.014.

Zabihi-Hesari, Alireza, Saeed Ansari-Rad, Farzad A. Shirazi, and Moosa Ayati. 2019. “Fault Detection and Diagnosis of a 12-Cylinder Trainset Diesel Engine Based on Vibration Signature Analysis and Neural Network.” *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 233(6):1910–23. doi: 10.1177/0954406218778313.