WILEY | Hindawi

Research Article

Vehicle Remote Health Monitoring and Prognostic Maintenance System

Uferah Shafi,[1] Asad Safi ⓘ,[1] Ahmad Raza Shahid,[1] Sheikh Ziauddin ⓘ,[1] and Muhammad Qaiser Saleem[2]

[1]COMSATS Institute of Information Technology, Islamabad 44000, Pakistan
[2]College of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia

Correspondence should be addressed to Asad Safi; asad.safi@gmail.com

In many industries inclusive of automotive vehicle industry, predictive maintenance has become more important. It is hard to diagnose failure in advance in the vehicle industry because of the limited availability of sensors and some of the designing exertions. However with the great development in automotive industry, it looks feasible today to analyze sensor's data along with machine learning techniques for failure prediction. In this article, an approach is presented for fault prediction of four main subsystems of vehicle, fuel system, ignition system, exhaust system, and cooling system. Sensor is collected when vehicle is on the move, both in faulty condition (when any failure in specific system has occurred) and in normal condition. The data is transmitted to the server which analyzes the data. Interesting patterns are learned using four classifiers, Decision Tree, Support Vector Machine, *K* Nearest Neighbor, and Random Forest. These patterns are later used to detect future failures in other vehicles which show the similar behavior. The approach is produced with the end goal of expanding vehicle up-time and was demonstrated on 70 vehicles of Toyota Corolla type. Accuracy comparison of all classifiers is performed on the basis of Receiver Operating Characteristics (ROC) curves.

1. Introduction

Vehicle systems are complex both in hardware and software so their maintenance is challenging. Maintenance strategy being used in vehicle industry is normally reactive that results in reduction of lifetime of vehicle and also loss of money. Predictive maintenance is required on this stage to overcome these issues. It is reported by European Commission that there will be 50% increment in transport vehicles within 20 years [1]. It will require effective strategies to keep up the vehicle performance. Vehicles having very complex structure need an effective maintenance strategy. Three types of maintenance strategies are being used in vehicle industry, predictive maintenance, corrective maintenance, and preventive maintenance. Preventive maintenance is performed after a fault has occurred. It is used for infrequent failures when the repair is extremely costly. Preventive maintenance is the common practice in the vehicle industry, where vehicle parts are upgraded occasionally. In contrast to preventive and corrective maintenance, in predictive maintenance [2], current condition of system/vehicle is analyzed to predict what is probably going to fail.

Vehicle has a very complex mechatronic structure consisting of subsystems, for example, gearbox, engine, and brakes [3]. Normally any subsystem comprises electromechanical processes, actuators, and sensors. The sensors and actuators are associated and controlled with an ECU (Engine Control Unit) which manages and screens the procedure. It is additionally associated with CAN (Controller Area Network) through which the distinctive subsystems and the driver communicate with each other. A high level diagnostic protocol is needed to communicate with Engine Control Unit (ECU). Two well-known protocols are OBD2 and UDS [4]. OBD2 alludes to a vehicle's self-indicative and reporting ability. On-Board Diagnostic (OBD) frameworks give the vehicle owner or a repair professional access to condition of current data of different vehicle subsystems while UDS does provide all details. System's current condition is evaluated by diagnostic

and prognostic processes. Diagnostics is concerned with current state of any subsystem whereas prognostic is related to the future state of subsystem [5].

There are serious challenges when we deal with prognostic maintenance. Prognostic maintenance copes with on-board data. Development cost of on-board diagnostic is limited in vehicle, which results in limited number of sensors. These sensors produce thousands of signals or data streams when vehicle is on the move. These signals are continuously sent to mobile or laptop attached with vehicle via wireless communication. It needs huge storage capacity which results in high cost. Therefore such systems are not implemented. Still there is tremendous increase in research on vehicle diagnostic during last decade. A detail discussion of predictive maintenance of vehicle industry is presented by Prytz [1] in which comprehensive analysis of how machine learning approaches are being used in automotive industry for fault prediction is also discussed. "Consensus self-organized models for fault detection (COSMO)" was presented [6] in which sensor's data is used. Heavy trucks and city bus were used in experiments. COSMO is presented to increase lifetime of vehicle. Another approach for vehicle's compressor fault prediction was reviewed [7] in which logged on-board data was used. Approach is demonstrated on Volvo trucks [8]; an approach was presented to predict need for repair in air compressor in buses and trucks. Random Forest was applied on logged on-board data for prediction.

Due to increasing complexity of vehicles, industry's focus has moved toward automated data analysis. In the meantime, with the minimal cost of wireless communication and increasing trend of android applications, it has become feasible to analyze on-board data for fault prediction. Naryal et al. [9] used android phone application, on-board vehicle diagnostic system for vehicle's health monitoring where driver was notified in case of any alarming conditions.

For each approach mentioned above, there are many options to perform diagnostic and prognostic maintenance. We propose complete infrastructure of vehicle remote health monitoring and prognostic maintenance system using data driven methodology which is based on real time data collected when vehicle is on move. Toyota Corolla model 2010 vehicles have been used for data generation. Toyota Corolla is series of cars developed by Corolla. Corolla started to manufacture these cars in 1966 and became best seller in 1997. Remote health monitoring refers to monitoring the working of different systems of vehicles remotely and prognostic refers to predicting fault in advance which is discussed in detail in Section 3. Our twofold contributions are fault prediction in four main systems of vehicle and providing real time vehicle monitoring and prognostic maintenance system. This paper presents an approach using sensor data and machine learning algorithms along with smart phone applications to completely analyze subsystems and provide fault prediction in subsystems. One great advantage of the proposed system could be in latest project of vehicle industry including autonomous car where every system in car is automatic; VMMS can help in automatic fault prediction of vehicle systems.

The rest of the paper is structured as Section 2 contains some related work; after that the proposed system architecture is presented in Section 3, experiments are in Section 4, process of feature selection is presented in Section 5, results are discussed in Section 6, and Section 7 presents conclusion and future work.

2. Related Work

Machine learning approaches are being used in vehicle industry since last decade to improve vehicle up-time. The existing technologies being used in vehicle industry include machine learning approaches, soft computing, and on-board data analysis. A new algorithm named "Sequential Pattern Mining Algorithm" was proposed [10] in which proposed algorithm learned patterns from warranty data of vehicles and the learned patterns are converted to rule based expert system. Choudhary et al. presented an overview of application being used in manufacturing industry using machine learning approaches including classification, clustering, and prediction [11]. Then with development in manufacturing of vehicle, trend moved from diagnostics of vehicles to prognostic using sensor data as large number of sensors were being implemented in vehicles. Using these sensor's data, a new model for fault prediction "Consensus Self-Organized Model for Fault Prediction" was presented Byttner et al. [6] in which model learned interesting relationships between sensor's data in vehicles and also in small mechatronic systems. Prytz et al. [12] presented an unsupervised method to find relations between sensor's data in two rounds; in first round, good relations were found by MSE (Mean Square Error); then in second round LASSO (Least Absolute Shrinkage and Selection Operator) and least square error were used to determine model parameters. Alzghoul et al. [13] used linear regression methods for fault prediction rather than classification and results showed that regression performed better than classification. Vehicle data bases containing vehicle's repairing records, including sensor's data, are normally imbalanced. A comprehensive technique "BRACID" (Bottom-up Induction of Rules And Cases for Imbalanced Data) to deal with imbalanced data in order to learn rules was presented by [14]; experiments and results showed that new algorithm performed much better as compared to classifiers including C4.5, PART, RIPPER, and CN2 in case of imbalanced data. Another techniques was presented [8] to learn classifiers from imbalanced data.

Rodger [15] proposed a vehicle health maintenance system using Kalman model; sensor data was used for fault prediction; moreover engine abnormal behavior was also observed by anomaly detection. Then in the same year another vehicle maintenance system was proposed for diagnostic and remote prognostic of vehicles based on least squares support vector machine classifier; the system promoted the use of smart phones in automotive industry. Remote maintenance of vehicle system is strongly based on vehicular communication network. Lu et al. [16] demonstrated vehicle ad hoc networks in detail and found parameters for communication. In preceding year, [17] presented a comprehensive analysis of failure prediction of compressor. Sensor data was used and also challenges and problems have been discussed in detail. Data was collected from logged

vehicle data base which maintains maintenance record of all vehicles visiting Volvo workshop. Random Forest, KNN, and C5.0 were used for fault prediction of compressor failure. Kargupta et al. [18] presented a vehicle monitoring system in their article. The system monitors driver activity and the status of the car engine. Smart phones were used for communication between vehicle and back-end server using wireless communication.

Since traffic is getting crowded day by day, accident danger increases. To overcome this situation, intelligent systems are being used with new technologies. Türker and Kutlu [19] presented an overview of existing systems using OBD2 tools and systems using communication through OBD2 interface. A comprehensive analysis of vehicle maintenance used unsupervised and supervised techniques with the help of telematics gateways which enables the vehicles to communicate with back-end server [1]. On-board and off-board data was used for vehicle fault diagnostic and prognostics. Then a novel data cloud service in Internet of things for vehicle was proposed [20] in which two data mining algorithms were used including Naive Bayes and Logistic Regression for warranty analysis using warranty data record of vehicles.

Sun presented multisensor fusion technique [21] to monitor vehicle's health using oil data and vibration signals. Schmalzel et al. presented a comprehensive discussion on how smart sensors could be used for health monitoring of a system and to diagnose a problem [22]. Another vehicle health monitoring system was presented by Murakami et al. using data base, data distribution network, and a communication system [23]. One more system for health monitoring was proposed by Ng et al. that was demonstrated on a passenger vehicle [24]. System was able to detect any problem or fault in sensors or actuators.

There are cost constraints when sensor data is being used as large memory space and processor speed are required. One solution for data reduction [25] was presented by Choi et al. using different machine learning techniques for engine. Another prognostic health monitoring system [26] was presented by Cole et al. using sensor data in which agent based system was used. Using on-board data, another remote health monitoring system [27] was proposed by Zhang et al. in which vehicle no start condition was analyzed in detail. Chen and Zhu presented complete model of vehicle health monitoring system [28] which was demonstrated on electric wheel truck. Naryal and Kasliwal presented an in vehicle embedded system [9] for health monitoring to analyze the internal condition of vehicles components by using travelers information. That system was also able to predict future failures to avoid interruption in journey. To monitor health state of machines and technical vehicles, a remote diagnostic and monitoring system [29] was proposed by Manakov et al. An application development project [30] was presented by Ganesan and Mydhile to monitor vehicles health condition remotely using self-adaptive technique. Ruddle et al. presented a prognostic vehicle's health monitoring system [2] for electric vehicles using failure analysis. To increase safety and reliability, Baraldi et al. presented a model for prognostic and health monitoring [31] for electric vehicles. Hodge et al. presented a survey of wireless sensors network [32] in railway

industry. Another android application based vehicle health monitoring system [33] was presented by Babu et al. in which engine's condition, battery condition, and emission system were monitored, and driver of vehicle was notified about the condition of mentioned systems via android phone.

We propose a real time vehicle monitoring and fault prediction system. In VMMS main subsystems of vehicle, ignition system, exhaust system, fuel system, and cooling system are analyzed in detail. Sensor data has been used for fault prediction using machine learning techniques, Decision Tree, SVM, K-NN, and Random Forest. Data is generated using smart phone and OBD2 scanner when vehicle is on drive. Sensor data in the form of Diagnostic Trouble Codes (DTC) is transmitted to smart phone via Bluetooth and then sent to back-end server. Classification algorithms are applied and interesting patterns are learned which can cause any system to fail. User is notified in case of abnormal condition via push notification on smart phone and email notification. User or owner of vehicle can monitor current condition of vehicle remotely. In next section, architecture of proposed VMMS is discussed in detail.

3. Overview of VMMS Architecture

VMMS has three main layers as shown in Figure 1. In first layer, data is generated. An OBD2 scanner is connected with the vehicle through OBD2 port. A variety of these scanners are available in market; ELM 327 Bluetooth (ELM 327) is being used in the proposed system which is basically a microcontroller. This scanner behaves like a bridge between vehicle and portable device, that is, mobile or laptop which supports Bluetooth. There are tiny ICs (integrated circuits) generated for this communication between vehicle and portable device. All sensor's data in form of DTC are generated when vehicle is on the move and sent to smart phone. Data is continuously being generated and transmitted to smart phone via Bluetooth. Actually smart phone communicates with ECU via wireless connection which is used as a source to connect vehicle to back-end serve. The conceptual diagram of VMMS is shown in Figure 1.

Figure 1 shows complete flow diagram of proposed VMMS. In data processing layer, first step is feature selection in which data stream of DTC is filtered in feature selection process using experts suggestion. Then a PCA (Principle Component Analysis) is applied on data set for feature reduction. After that four classification algorithms are used in classification phase including Decision Tree, Random Forest, K-NN, and SVM. Interesting combinations of DTC or relationships are learned and further processing is done on server end. These results are stored on server for further derivation which is used for fault prediction and remote monitoring of vehicle.

In remote monitoring layer, the owner or concerned person of vehicle can monitor the current condition of vehicle remotely like fuel status, speed, and current position. The driver or the owner of vehicle is notified about the failure of any subsystem of vehicle through automatic notification.

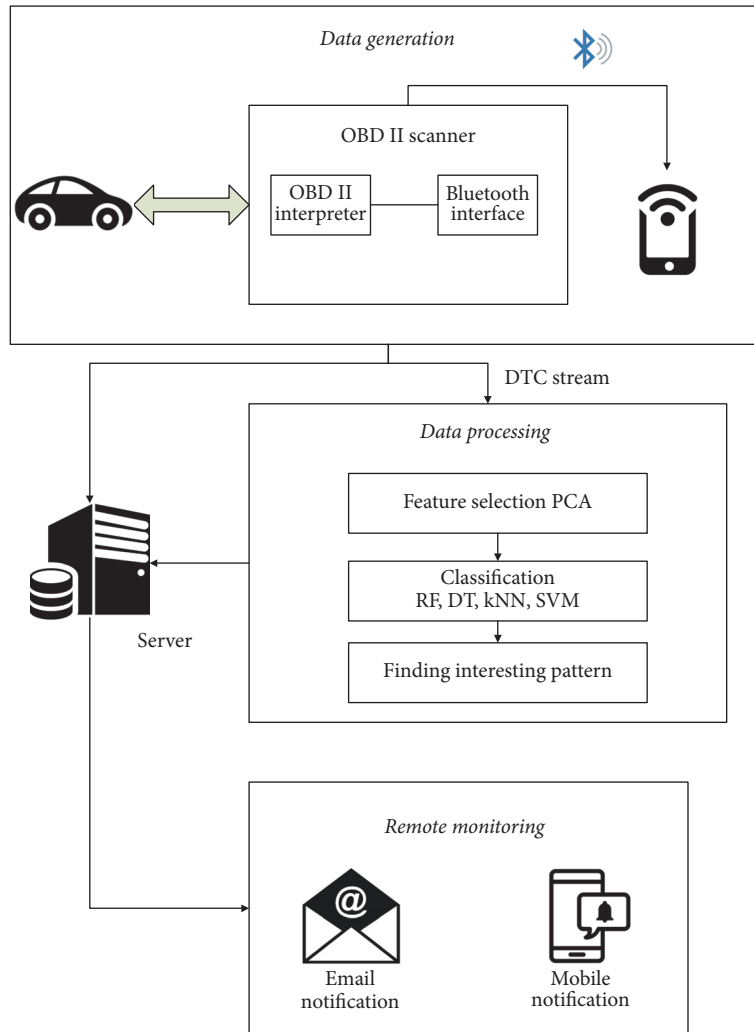The main advantages of the proposed system are as follows:

Figure 1: VMMS Architecture.

(i) A nontechnical person can get into the failure of any subsystem of vehicle as VMMS provides all information of vehicle status and condition on smart phone which is connected to vehicle.

(ii) VMMS focuses on the real life solutions using machine learning techniques which can save money and time as well.

(iii) Autonomous vehicle needs continuous stream of sensory input of functioning part of vehicle; VMMS prediction will guide autonomous vehicles decision system from failure of any system part.

(iv) Vehicle life time is increased as when owner/driver knows all about internal conditions of systems; then he can get some steps to get rid of any system failure

(v) Accident risk decreases with the awareness of structure of systems.

4. Data Generation

Vehicle four main systems are analyzed including ignition system, fuel system, exhaust system, and cooling system for data generation. Those systems are monitored by sophisticated computer controls. Sensors provide fault information to Engine Control Unit (ECU). ECU is a microcomputer which consists of lot of electronic components and circuits including many semiconductor devices. Its input device receives input in the form of electric signals. These signals come from many sensors located at different positions in engine. Its processing unit compares input data with data stored in memory. The memory unit contains basic information about how to operate engine. The output device pulses the electric signals of the solenoid type injection valve. ECU basic function in electric fuel injection system is to control pulse rate through injector, idle speed, ignition timing, and fuel pump. The ECU adjusts quickly to change the conditions by using programmed characteristics map stored in memory

unit. Aim of ECU is to maximize engine power with lowest amount of exhaust emission and lowest fuel consumption. OBD2 scanner tools download on-board trouble codes by communicating with ECU to determine which sensor is not helping (ECU). Then there is a serial bus communication protocol CAN (Controller Area Network) which is made by Boch in 1980s. It describes a standard for effective and trustworthy communication between controller, actuator, sensor, and ECU.

When vehicle is on drive, OBD2 scanning tool is connected with vehicle and DTCs generated on run time are being transmitted to the smart phone. ELM327, the scanning tool being used, can display more than 1500 values of sensor data. Toyota Corolla vehicles have been used in experiments. Data of around 70 vehicles has been collected, both normal case when there is no fault and other case when faults were there in different vehicle subsystems. A stream of DTC is produced by sensors with sampling frequency of 1 Hz when vehicle is on the move for each system under experiment. Each reading is taken as one example. Data set consists of 150 examples. DTC generated by sensors is considered as an attribute or feature. The feature value is set 1 if that particular DTC is generated and set as 0 otherwise. The output of system or class label is also in binary form. If system is on operation then output is set as 0 which means that vehicle is in safe condition. If fault occurred or system breaks down, then output class label is set as 1. So data set generated is completely binary in nature.

5. Feature Selection

Around 20 DTCs were produced in each reading. Nine features for each system have been selected carefully by experts by mutual understanding. Moreover PCA has been applied to prioritize the feature on the basis of variance which is basically square of standard deviation (s).

$$\text{var}(X) = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Xi - \overline{X})}{n-1}, \tag{1}$$

where (2) shows standard deviation and (3) shows mean $\overline{X}$:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{n-1}} \tag{2}$$

$$\overline{X} = \frac{\sum_{i=1}^{i} Xi}{n}. \tag{3}$$

The cut-off value of PCA is selected as 95%. Since the highest eigenvalues represent the most relevant components, a cut-off value C_{PCA} is chosen; the cut-off value C_{PCA} is chosen empirically from the data. We compared the selected featured by experts with results of PCA. Almost the same 9 features contained 95% of variation as selected by experts; rest of the 11 features contained 5% of variation. Therefore those 11 features have been excluded as they contained small information regarding failure of a specific subsystem as shown in the Figure 2.
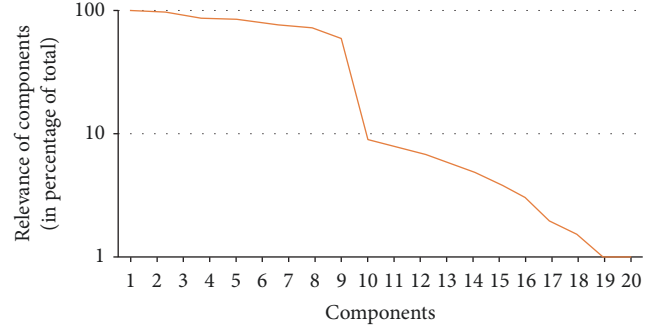

FIGURE 2: Representative example of the sorted PCA eigenvalue. The x-axis shows the values of the component as a percentage of the total in log scale.

Those classification algorithms are selected which perform good on binary data, that is, with high accuracy and minimum error rate.

Decision Tree: Basic algorithm for Decision Tree [34] is applied using Gini (gdi) Diversity Index (measure of node impurity) for splitting criteria calculated by

$$(\text{gdi}) = 1 - \sum_{i} p^2(i), \tag{4}$$

where $p(i)$ is probability that particular instance belongs to class Ci, that is, class 0 or class 1. Maximum number of splits is set to 10 to restrict tree size.

Support Vector Machine: SVM is optimization problem classifier [35]. In binary class problem, SVM draws a line that separates the instances of two classes and classify test instance by maximizing the margin between both sides of line. Radial Basis Function kernel (K) is selected for similarity measure.

$$K(x, x') = \exp\left(-\eta \left|x - x'\right|^2\right), \tag{5}$$

where is η a parameter to handle nonlinear classification.

K-Nearest Neighbor: kNN [36] is applied where Euclidean distance (d) measure is used to calculate similarity or distance and number of neighbor (k) is set to 5.

$$d(X, Y) = \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}. \tag{6}$$

Random Forest [37]. An ensemble learning method is applied to improve the accuracy, that is, Random Forest. Decision Tree algorithm is used; the maximum number of learners is set to 20. Data is sampled n-times with replacement; for each sample, a Decision Tree is learned. So, at the end, there are n Decision Trees. Test instance is classified on the basis of majority votes of learned classifiers (Decision Trees).

6. Results and Discussion

In this section, all experimental results of classifiers are being discussed in detail.

Four classifiers are selected including Decision Tree, Support Vector Machine, K-NN, and Bagging Tree (Random Forest). As the data set is binary in nature so these algorithms are selected as they perform good on binary data. These algorithms are implemented on each system including ignition system, exhaust system, fuel system, and cooling system. The performance of each algorithm on particular system is evaluated on the basis of accuracy, precision, recall, and $F1$ score measures. 10-fold cross validation has been used. Accuracy is the percentage of the total number of predictions that was found correct, whereas precision is true positive accuracy (tpa), $F1$ score is accuracy which is measured using precision and recall, and recall is true positive rate (tpr). According to [38, 39], the precision, $F1$ score, accuracy, and recall are calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

$$F1 \text{ Score} = \frac{2 * P * R}{(P + R)} \tag{7}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FN} + \text{FP})}.$$

ROC curve is basically a graphical representation which illustrates the performance of binary classifier. It is actually used to compare the accuracy of different classifiers.

Now, detailed discussion of classification results on each system is presented then the results of the proposed system VMMS are compared with base papers [4, 7].

6.1. Ignition System. Ignition system is very important as every system depends on it. Decision Tree, SVM, KNN, and Random Forest (Bagging Tree) are applied. All classifiers are trained on 150 readings. ROC, precision, $F1$ score, accuracy, and recall measures have been used to evaluate the classifier. Performance comparison between classifiers is performed using ROC curve as shown in Figure 3.

Performance comparison of all classifiers including Decision Tree, SVM, K-NN, and Random Forest is shown in Figure 3. Accuracy is measured by how much area is covered by ROC curve. As Figure 3 shows SVM covers highest area under the curve with 0.99% and Decision Tree covers lowest area under the curve as 0.73%. K-NN covers 0.82% area under curve whereas FR covers 0.73% area under curve. Results show that SVM performs much better as compared to other classifiers. Table 1 shows precision, recall, accuracy, and F1 measure values for each classifier.

SVM beats other classifiers with 96.6% accuracy, 0.94 precision, and 0.98 recall; DT shows poor performance with 72.5% accuracy, 0.73 precision, and 0.5 recall.

6.2. Fuel System. For failure prediction of fuel system, selected four classifiers are applied on data set including Decision Tree, SVM, K-NN, and Random Forest; results show that SVM performs much better as compared to other classifiers as shown in Table 2.
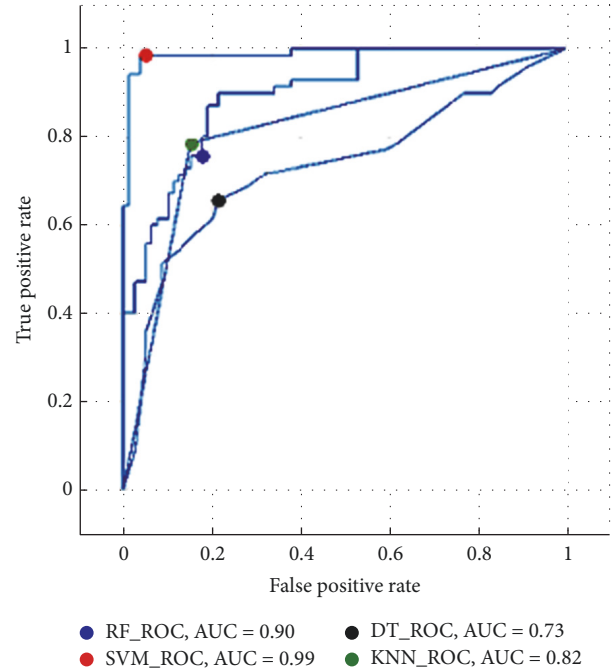

FIGURE 3: ROC comparison (ignition system).

TABLE 1: Ignition system VMMS.

Classifiers	Precision	Recall	Accuracy	$F1$ score
DT	0.73	0.5	72.5	0.68
SVM	0.94	0.98	96.6	0.96
kNN	0.82	0.78	81.9	0.77
RF	0.79	0.75	79.2	0.76

TABLE 2: Fuel system VMMS.

Classifiers	Precision	Recall	Accuracy	$F1$ score
DT	0.77	0.66	76.5	0.69
SVM	0.98	0.98	98.7	0.97
kNN	0.9	0.98	94.6	0.93
RF	0.86	0.90	90	0.87

SVM shows 98.7% accuracy with 0.98 precision and recall. DT shows 76.5% accuracy with 0.77 precision and 0.66 recall. In the same way K-NN and RF performance measures are given in Table 2.

Similarly for performance comparison of these classifiers, ROC curve comparison is presented in Figure 4.

SVM covers maximum area under curve which means SVM shows highest accuracy where Decision Tree shows poor performance with 0.77% area under curve. For fuel system, K-NN and RF perform much better with area under curve of 0.95% and 0.97%, respectively.

6.3. Exhaust System. Performance of each classifier is judged on the basis of accuracy, precision, recall, and $F1$ score which is shown in Table 3.
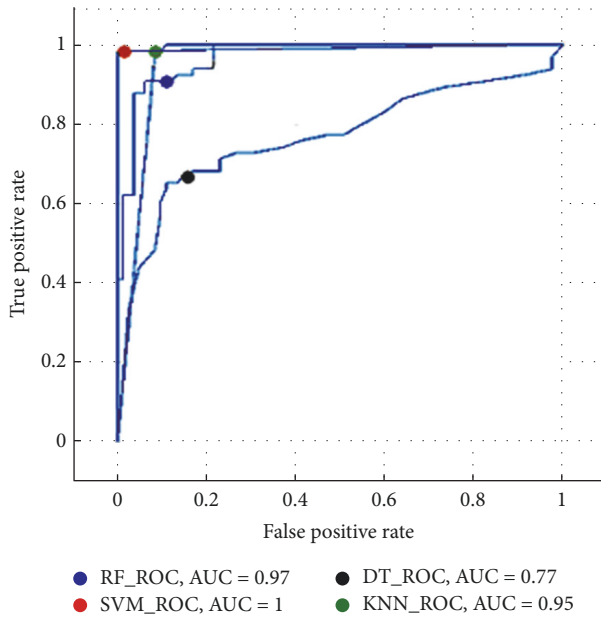
FIGURE 4: ROC comparison (fuel system).

- RF_ROC, AUC = 0.97
- SVM_ROC, AUC = 1
- DT_ROC, AUC = 0.77
- KNN_ROC, AUC = 0.95

FIGURE 5: ROC comparison (exhaust system).

- RF_ROC, AUC = 0.97
- SVM_ROC, AUC = 0.99
- DT_ROC, AUC = 0.76
- KNN_ROC, AUC = 0.90

TABLE 3: Exhaust system VMMS.

Classifiers	Precision	Recall	Accuracy	$F1$ score
DT	0.75	0.65	78.5	0.68
SVM	0.97	0.98	98	0.97
kNN	0.89	0.87	89.9	0.87
RF	0.84	0.90	76.6	0.86

TABLE 4: Cooling system VMMS.

Classifiers	Precision	Recall	Accuracy	$F1$ score
DT	0.72	0.65	75.8	0.68
SVM	0.96	0.95	96.6	0.95
kNN	0.92	0.95	94.6	0.93
RF	0.87	0.87	89.3	0.86

SVM shows good performance with accuracy 0.97%; K-NN and RF show approximately the same accuracy with 0.87% and 0.86% whereas DT shows 78.5%.

Accuracy comparison of all classifiers on the basis of ROC curve is shown in Figure 5.

In the case of exhaust system, RF and SVM show high accuracy with area under curve including 0.97% and 0.99%. K-NN shows 0.90% area under curve and DT shows poor accuracy with lowest area under curve with 0.76%.

6.4. Cooling System. Same algorithms including Decision Tree, SVM, KNN, and Random Forest have been applied. Precision, $F1$ measure, recall, and accuracy values for cooling system are shown in Table 4, which shows that SVM performs much better as compared to other classifiers.

In the case of cooling system, SVM and K-NN show highest accuracy with 96.6% and 94.6%, respectively. DT shows poor accuracy with 0.72%.
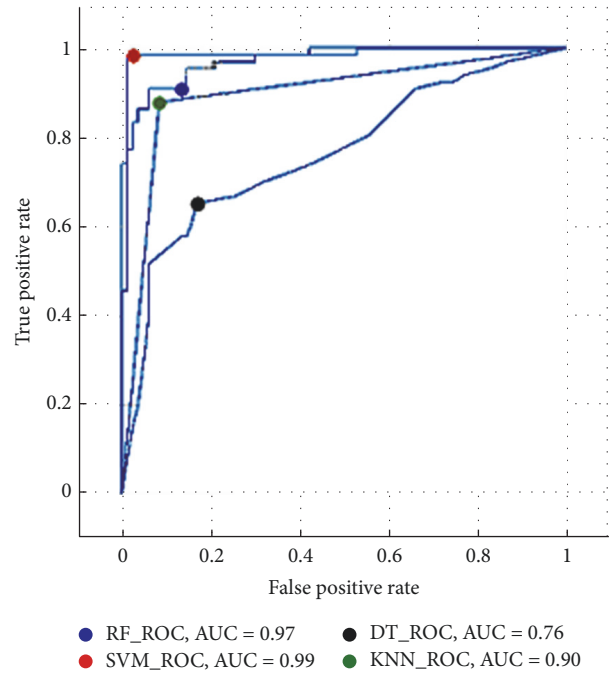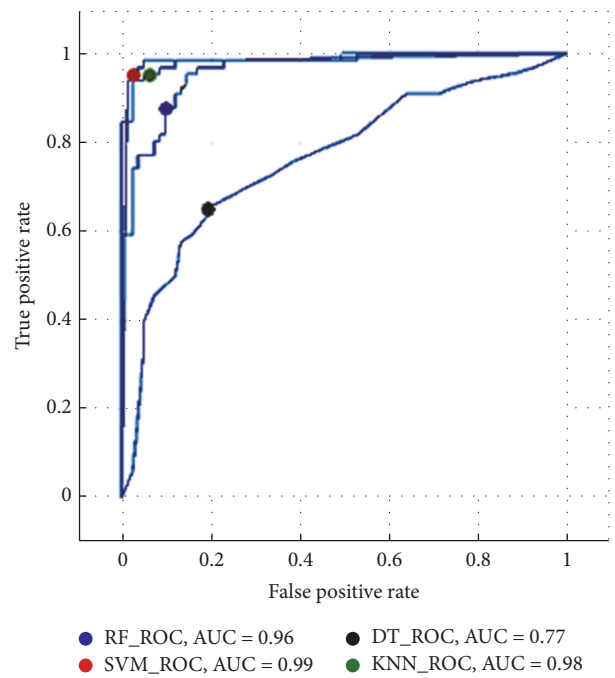

FIGURE 6: ROC comparison (cooling system).

- RF_ROC, AUC = 0.96
- SVM_ROC, AUC = 0.99
- DT_ROC, AUC = 0.77
- KNN_ROC, AUC = 0.98

ROC curves of each algorithm are shown in Figure 6 to compare which classifier performs best.

Figure 6 shows that SVM shows highest accuracy with area under curve 0.99% whereas Decision Tree shows poor accuracy with area under curve 0.77%.

Figures 3, 4, 5, and 6 show that SVM hits highest accuracy of approximately 98% whereas Decision Tree shows poor
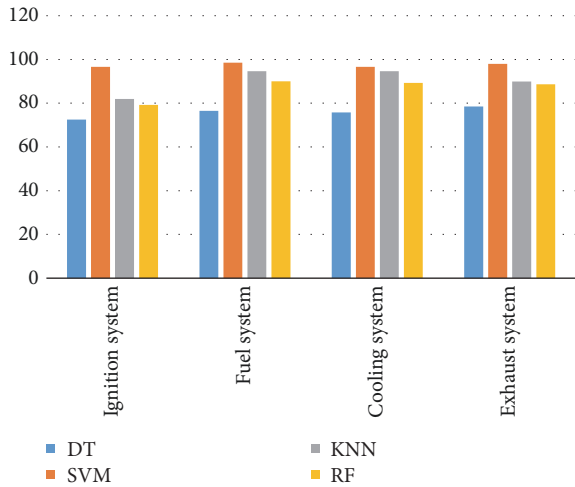
FIGURE 7: Accuracy comparison.

TABLE 5: Comparison with existing systems.

Systems VMMS	DT%	SVM%	kNN%	RF%
Ignition system	72.5	96.6	81.9	79.2
Fuel system	76.5	98.5	94.6	90.0
Exhaust system	78.5	98	89.9	88.6
Cooling system	75.8	96.6	94.6	89.3

performance. According to the survey of classifiers [40], the main reason for outstanding performance of SVM is that decision surface is drawn based on boundary cases. SVM does not consider entire data. SVM performs very good on small observations. In current situation, data for 150 experiments is collected. That is why SVM achieved highest accuracy. In contrast to SVM, Decision Tree considers all data set and classify instances in step wise process. A small change in node selection can results in big change in result. The other flaw of Decision Tree is that it strongly depends on training data set, so it often shows poor performance on test data set. So these are the reasons that can cause poor performance in the case of Decision Tree. K-NN and RF neither showed outstanding performance nor showed poor performance. On average K-NN and RF showed good performance.

Now accuracy comparison with all systems being discussed in this paper is presented in Table 5 which shows the accuracy comparison of four classifiers including Decision Tree/C5.0, Support vector machine, K Nearest Neighbor, and Random Forest/Bagging Tree on different systems of vehicles.

The accuracy of these four classifiers is compared with proposed VMMS ignition system, VMMS fuel system, VMMS exhaust system, and VMMS cooling system graphically in Figure 7 which shows graphical interpretation of accuracy comparison.

Graphical interpretation of Table 5 is shown in Figure 7.

In this section we discussed classification results of each system of vehicle including ignition system, fuel system, exhaust system, and cooling system and accuracy comparison

of existing systems. Next section contains conclusion and future work.

7. Conclusion and Future Work

With the increasing trend of smart phones and wireless communication, it has become feasible to use these technologies for real time solutions. Despite limited resources, these technologies are being used along with machine learning approaches to solve big problems in automotive industry. A novel vehicle monitoring and fault predicting system is presented in this paper including VMMS. Four classifiers including Decision Tree, SVM, RF, and K-NN have been used for fault prediction. Four main systems of vehicles have been considered. The main objective of the proposed system is to reduce the fault frequency of systems in vehicle.

Our first contribution to this paper is data generation and feature selection. Sensor data of Corolla cars manufactured by Toyota has been used. We selected those distinctive sensors that can cause a system to break down. The second contribution is to propose a user friendly vehicle fault prediction and vehicle remote health monitoring system.

There are several ways to extend current results. A noticeable thing is to find the ways of refining classification performance. Four classification algorithms have been used, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbor. The accuracy can also be discussed by applying other prediction techniques and working more on data set. All data that we have collected comes from 70 vehicles of Toyota Corolla type. In the future, we intend to work on various types of vehicles to gather data. We will continue our effort in this field to improve results and to explore more systems in vehicles like CAN, and so on.

Acronyms

OBD:	On-board diagnostic
DTC:	Diagnostic Trouble Code
ROC:	Receiver Operating Characteristics
AUC:	Area under the curve
ECU:	Engine Control Unit
CAN:	Controller Area Network
PCA:	Principle Component Analysis
DT:	Decision Tree
SVM:	Support Vector Machine
K-NN:	K-Nearest Neighbor
RF:	Random Forest
NA:	Not available
TP:	True positive
TN:	True negative
FP:	False positive
FN:	False negative
P:	Precision
R:	Recall.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] R. Prytz, *Machine Learning Methods for Vehicle Predictive Maintenance Using Off-Board And On-Board Data Dissertation*, Halmstad University Press, 2014.

[2] A. R. Ruddle, A. Galarza, B. Sedano, I. Unanue, I. Ibarra, and L. Low, "Safety and failure analysis of electrical powertrain for fully electric vehicles and the development of a prognostic health monitoring system," in *Proceedings of the IET Hybrid and Electric Vehicles Conference 2013, HEVC 2013*, pp. 1–6, IET, London, UK, November 2013.

[3] J. Jeong, N. Kim, W. Lim, Y. Park, S. W. Cha, and M. E. Jang, "Optimization of power management among an engine, battery and ultra-capacitor for a series HEV: a dynamic programming application," *International Journal of Automotive Technology*, vol. 18, no. 5, pp. 891–900, 2017.

[4] M. A. Taie, M. Diab, and M. Elhelw, "Remote prognosis, diagnosis and maintenance for automotive architecture based on least squares support vector machine and multiple classifiers," in *Proceedings of the 2012 4th International Congress on Ultra Modern Telecommunications and Control Systems, ICUMT 2012*, pp. 128–134, Russia, October 2012.

[5] G. Shi, P. Dong, H. Q. Sun, Y. Liu, and Y. X. Cheng, "Adaptive control of the shifting process in automatic transmissions," *International Journal of Automotive Technology*, vol. 18, no. 1, pp. 179–194, 2017.

[6] S. Byttner, T. Rögnvaldsson, and M. Svensson, "Consensus self-organized models for fault detection (COSMO)," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 5, pp. 833–839, 2011.

[7] R. Prytz, S. Nowaczyk, T. Rögnvaldsson, and S. Byttner, "Analysis of truck compressor failures based on logged vehicle data," in *Proceedings of the International Conference on Data Mining (DMIN)*, p. 1, 2013, The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

[8] R. Prytz, S. Nowaczyk, T. Rögnvaldsson, and S. Byttner, "Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 139–150, 2015.

[9] E. Naryal and P. Kasliwal, "Real time vehicle health monitoring and driver information display system based on CAN and Android," *International Journal of Advance Foundation and Research in Computer*, vol. 1, no. 11, pp. 76–84, 2014.

[10] J. Buddhakulsomsiri and A. Zakarian, "Sequential pattern mining algorithm for automotive warranty data," *Computers & Industrial Engineering*, vol. 57, no. 1, pp. 137–147, 2009.

[11] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: a review based on the kind of knowledge," *Journal of Intelligent Manufacturing*, vol. 20, no. 5, pp. 501–521, 2009.

[12] R. Prytz, S. Nowaczyk, and S. Byttner, "Towards relation discovery for diagnostics," in *Proceedings of the 1st International Workshop on Data Mining for Service and Maintenance, KDD4Service 2011—Held in Conjunction with SIGKDD'11*, pp. 23–27, USA, August 2011.

[13] A. Alzghoul, M. Löfstrand, and B. Backe, "Data stream forecasting for system fault prediction," *Computers & Industrial Engineering*, vol. 62, no. 4, pp. 972–978, 2012.

[14] K. Napierala and J. Stefanowski, "BRACID: a comprehensive approach to learning rules from imbalanced data," *Journal of Intelligent Information Systems*, vol. 39, no. 2, pp. 335–373, 2012.

[15] J. A. Rodger, "Toward reducing failure risk in an integrated vehicle health maintenance system: a fuzzy multi-sensor data fusion Kalman filter approach for IVHMS," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9821–9836, 2012.

[16] W. Lu, L. D. Han, and C. R. Cherry, "Evaluation of vehicular communication networks in a car sharing system," *International Journal of Intelligent Transportation Systems Research*, vol. 11, no. 3, pp. 113–119, 2013.

[17] J. Stefanowski, "Overlapping, rare examples and class decomposition in learning classifiers from imbalanced data," *Smart Innovation, Systems and Technologies*, vol. 13, pp. 277–306, 2013, Emerging Paradigms in Machine Learning.

[18] H. Kargupta, R. Bhargava, K. Liu et al., "VEDAS: a mobile and distributed data stream mining system for real-time vehicle monitoring," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, pp. 300–311, Philadelphia, PA, USA, 2004, Society for Industrial and Applied Mathematics.

[19] G. F. Türker and A. Kutlu, "Evaluation of car accident prevention systems through onboard diagnostic," *Global Journal on Technology*, vol. 7, pp. 01–06, 2015.

[20] W. He, G. Yan, and L. D. Xu, "Developing vehicular data cloud services in the IoT environment," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1587–1595, 2014.

[21] Q. Sun, "Sensor fusion for vehicle health monitoring and degradation detection," in *Proceedings of the 5th International Conference on Information Fusion, FUSION 2002*, pp. 1422–1427, USA, July 2002.

[22] J. Schmalzel, F. Figueroa, J. Morris, S. Mandayam, and R. Polikar, "An architecture for intelligent systems based on smart sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1612–1616, 2005.

[23] T. Murakami, T. Saigo, Y. Ohkura, Y. Okawa, and T. Taninaga, "Development of vehicle health monitoring system (VHMS/WebCARE) for large-sized construction machine," *Komatsu's Technical Report*, vol. 48, no. 150, pp. 15–21, 2002.

[24] H. K. Ng, R. H. Chen, and J. L. Speyer, "A vehicle health monitoring system evaluated experimentally on a passenger vehicle," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 5, pp. 854–870, 2006.

[25] K. Choi, J. Luo, K. R. Pattipati, S. M. Namburu, L. Qiao, and S. Chigusa, "Data reduction techniques for intelligent fault diagnosis in automotive systems," in *Proceedings of the 2006 IEEE AUTOTESTCON - IEEE Systems Readiness Technology Conference*, pp. 66–72, USA, September 2006.

[26] I. S. Cole, P. A. Corrigan, W. Ganther et al., "Development of a sensor-based learning approach to prognostics in intelligent vehicle health monitoring," in *Proceedings of the 2008 International Conference on Prognostics and Health Management, PHM 2008*, USA, October 2008.

[27] Y. Zhang, M. Salman, H. S. Subramania et al., "Remote vehicle state of health monitoring and its application to vehicle no-start prediction," in *Proceedings of the IEEE AUTOTESTCON 2009 - Systems Readiness Technology Conference: Mission Assurance Through Advanced ATE*, pp. 88–93, USA, September 2009.

[28] S. Chen and M. Zhu, "Design and realization of health monitoring system on electric wheel truck," in *Proceedings of the 2010*

International Symposium on Intelligence Information Processing and Trusted Computing, IPTC 2010, pp. 51–54, China, October 2010.

[29] A. L. Manakov, A. A. Igumnov, and S. A. Kolarzh, "Monitoring technical state of transportation vehicles and production machines," *Journal of Mining Science*, vol. 49, no. 4, pp. 630–636, 2013.

[30] R. Ganesan and S. Mydhile, "Design and development of remote vehicle health monitoring system using context aware web services," in *Proceedings of the 2013 IEEE Conference on Information and Communication Technologies, ICT 2013*, pp. 760–764, ind, April 2013.

[31] P. Baraldi, F. Di Maio, M. Rigamonti et al., "A Procedure for Practical Prog- nostics and Health Monitoring of Fully Electric Vehicles for Enhanced Safety and Reliability," in *Proceedings of the 5th IET Hybrid and Electric Vehicles Conference (HEVC 2014)*, London, UK, 2014.

[32] V. J. Hodge, S. O'Keefe, M. Weeks, and A. Moulds, "Wireless sensor networks for condition monitoring in the railway industry: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1088–1106, 2015.

[33] M. S. Babu, T. A. Anirudh, and P. Jayashree, "Fuzzy system based vehicle health monitoring and performance calibration," in *Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, pp. 2549–2554, India, March 2016.

[34] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, Mass, USA, 1984.

[35] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.

[36] I. Mani and I. Zhang, "kNN approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of Workshop on Learning from Imbalanced Datasets*, p. 126, 2003.

[37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[38] D. M. Powers, "Evaluation, from precision, recall and F-measure to ROC, informedness," *Markedness and Correlation*, 2011.

[39] S. Russell and P. Norvig, "A modern approach," in *Artificial Intelligence*, vol. 25, Prentice Hall, Englewood Cliffs, NJ, USA, 1995.

[40] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, 2006.