Engine Misfire Detection with Pervasive Mobile Audio

Joshua Siegel[(⊠)], Sumeet Kumar, Isaac Ehrenberg, and Sanjay Sarma

Massachusetts Institute of Technology, Cambridge, MA 02139, USA
{j_siegel,sumeetkr,yitzi,sesarma}@mit.edu

Abstract. We address the problem of detecting whether an engine is misfiring by using machine learning techniques on transformed audio data collected from a smartphone. We recorded audio samples in an uncontrolled environment and extracted Fourier, Wavelet and Mel-frequency Cepstrum features from normal and abnormal engines. We then implemented Fisher Score and Relief Score based variable ranking to obtain an informative reduced feature set for training and testing classification algorithms. Using this feature set, we were able to obtain a model accuracy of over 99 % using a linear SVM applied to outsample data. This application of machine learning to vehicle subsystem monitoring simplifies traditional engine diagnostics, aiding vehicle owners in the maintenance process and opening up new avenues for pervasive mobile sensing and automotive diagnostics.

Keywords: Pervasive sensing · Mobile phones · Sound classification · Audio processing · Fault detection · Machine learning

1 Introduction

People spend more time in their cars than ever before, and with growing miles traveled [25], hours spent in traffic [18], and an aging vehicle fleet in the United States and around the world [7], vehicle maintenance has become an increasingly critical part of vehicle ownership. Proactive or rapid-response maintenance saves significant cost over the life of a vehicle and reduces the likelihood of an unplanned breakdown. Anticipatory maintenance can further alleviate reliability concerns and increase the overall satisfaction of vehicle owners and operators through reduced fuel consumption, emissions, and improved comfort. For these reasons, the consumer-facing diagnostic market for vehicles has grown to include products intended to help vehicle owners maintain and supervise the operation of their vehicles without the assistance of a mechanic.

At the core of any vehicle's maintenance requirement is the engine, responsible for efficient and reliable propulsion. Automotive internal combustion engines require only three "ingredients" to run: a supply of fuel, intake air, and ignition sparks. Delivery of one or more of these elements can fail, as is the case when an air filter or fuel injector clogs or when an ignition coil is damaged. One common engine fault results from a weak or non-existent spark, causing the fuel

in a cylinder to fail to combust. With one or more cylinders failing to explode and generate motive force, fuel efficiency and power output drop, with the engine operation increasing in noise, vibration, and harshness. This fault, called a "misfire," results in engine wear and leads to hesitation upon acceleration. A weak spark may be the result of neglected maintenance, such as fouled spark plugs, or component failure, such as an intermittently connected plug wire or an ignition coil pack stressed from powering improperly gapped spark plugs.

Per a 2011 CarMD "Vehicle Health Index" [2], misfires are severe faults and the most commonly occurring vehicle failure, representing 13.8 % of reported problems. Beyond the cost of damage resulting from inaction, misfires have the potential to incur significant additional fuel costs resulting from inefficient or incomplete combustion.

In modern vehicles, computer systems monitor combustion, misfires, and other emission-related functions through a system called "On Board Diagnostics" (OBD) [17]. While OBD systems are capable of detecting a misfire, they are slow to react, rely on proprietary and non-standard algorithms, and necessitate the use of a specialized interface device to provide human-readable information. In a survey we conducted of 15 drivers who had recent, active check-engine lights, we determined that owners left problems unaddressed an average of 3,500 miles [20]. Though OBD tools are available, we determined that they are underutilized by the average vehicle owner.

To better enable preventative maintenance, it is desirable to instead detect these faults passively, more reliably, and without specialized equipment, applying sensing from devices such as mobile phones and allowing location- and orientation-independent analysis. This would remove the barrier to entry posed by requiring a dedicated code-reading device and enable pervasive sensing to allow drivers to monitor the health of their vehicles with increasing frequency at no additional cost. Through improved early detection, the source of the misfire can be addressed easily and inexpensively with the replacement of a spark plug, wire, or ignition coil, before the failure takes a more costly toll on other components like the catalytic converter due to long-term rich fuel trim.

A concurrent proliferation in mobile devices, along with recent advances in sensing and computation, has made pervasive sensing a valuable field for exploration. The use of mobile phones as "automotive tricorders" capable of non-invasively detecting vehicle condition will encourage drivers to take an active role in vehicle maintenance through improved ease-of-use and widespread adoptability relative to current diagnostic offerings. Passive sensing will allow a shift from today's paradigm of reactive repair to one of proactive maintenance, with this technique having been used successfully for passive monitoring of wheels and tires [21,22].

In this paper, we show that pervasive sensing may be used to differentiate normally operating engines from those operating with misfires. Because we lack a robust physical model describing misfire phenomena, we apply machine learning techniques to uncontrolled data collection and demonstrate an approach to misfire detection making use of extensive feature generation and set reduction to

improve classification without a physically-derived hypothesis. We demonstrate that a mobile device may be used to generate data, create a set of features, reduce the size of that set, and apply machine learning to classify accurately and efficiently based on the reduced set.

This paper covers topics ranging from data collection to feature generation to classification. In Sect. 2, we consider prior art and how our method differs from the in-situ and externally sensed solutions before it, illustrating the opportunity space and motivating our work. Section 3 describes our approach to data generation, and how we minimize experimental setup in favor of more naturalistic and representative data collection capable of more easily translating to a consumer-friendly application. Section 4 explores the algorithms we use to generate a comprehensive feature vector. Our approach applies exhaustive feature generation because we have no prior art to distinguish what might be important to classify misfires from normally operating engines. We further discuss our approach to reducing feature set size using feature ranking techniques to facilitate lower computational and other resource overheads. In Subsect. 4.3, we briefly discuss the various classification algorithms we implemented and their relative merits, drawbacks, and efficacy. We conclude in Sect. 5 and show 99 % classification accuracy with 50 % outsample data, before Sect. 6 which discusses plans for future work in this area.

2   Prior Art

Engine misfires have been detected in a variety of ways. Under normal operating conditions, the crankshaft rotates through a fixed angular displacement between every cylinder firing attempt. A misfire detectably alters the precession of the crankshaft which is sensed by a crankshaft position sensor. Measuring a series of unexpected angular measurements within a time window prompts the illumination of a check engine light indicating that the engine is operating outside of specifications and malfunctioning. The use of an OBD scan tool may reveal which cylinder or cylinders are misfiring, but this information is of uncertain provenance and dubious value due to the use of proprietary classification schemes [14,19]. Some direct-sensing alternatives to crankshaft position-based detection include sampling of the instantaneous exhaust gas pressure, measuring ionization current in the combustion chamber, or installing other sensors within [5,27] or outside the combustion chamber [15,26].

Other diagnostics have demonstrated the capacity to identify misfires through audio signal processing. Aside from less obviously discernible symptoms like increased fuel consumption or visual indications like an oily or white residue on the tip of the spark plug, misfires have a characteristic audible "pop" and cause the engine to vibrate as though it is unbalanced or otherwise "missing a beat". The sound emanating from an abnormally-firing engine can be captured at a distance by a microphone and analyzed in both the time and frequency domains for patterns indicative of cylinder misfires. Auto mechanics have long employed a form of auditory diagnosis, listening to engines and easily determining the

presence of a misfire. The fact that physical models of the sound and vibration profiles produced by an engine misfire are complex, yet experienced mechanics can classify a firing abnormality by ear, lends credence to the idea that a machine learning approach to detection may be tenable.

Researchers have applied this sort of classification technique successfully. To acquire the audio signals, Dandare [4] and Sujono [23] made use of dedicated recording equipment to analyze the sound from automotive internal combustion engines in a laboratory environment. Engines were recorded during normal operation, as well as in the presence of different faults including cylinder misfire. In Dandare, an Artificial Neural Network classified faults with accuracies ranging from 85–95% overall. Kabiri and Ghaderi [8,9] introduced noise into their misfire measurements of over 300 single cylinder engines by moving outside the laboratory and into a garage. Principal Component Analysis and correlation-based feature selection in the time and frequency domains achieved an accuracy between 70 % and 85 % for these vehicles. Anami made similar recordings of motorcycles [1] to aid mechanics in the rapid classification of healthy versus faulty. Several hundred motorcycle engines were recorded from a distance of half a meter, with wavelet-based machine learning techniques distinguishing not only between healthy and faulty motorcycles, but also the category of fault present. This included, for example, whether the fault was in the engine or exhaust system. Experienced mechanics provided ground truth, with the classification system reporting > 85% accuracy relative to these uncertain reference values.

With the proliferation of smartphones among the car owning public, researchers have considered how these devices can be used to aid in vehicle diagnostics and more specifically engine misfire detection. Using the smartphone at the center of a remote maintenance system, Tse [24] installed sensors including accelerometers and laser encoders within a test vehicle. When a misfire or other engine events were detected, a message was sent via smartphone to inform the user. In Navea [16], the smart phone itself was used as the data collection and processing device, and was held 30 cm above the engine cover to record sounds of the engine and drive belt during startup, while idling, and at 1000 RPM. Thirty-five Honda Civics were used and recordings were taken at various locations and ambient conditions as input data for Fast-Fourier Transformed data based classification. Startup issues relating to the car battery, fuel supply and timing were recognized 100 % of the time, while a normal engine at idle or 1000 RPM was identified with a 33 % false positive rate. Pulley bearing defects or belt slips were properly diagnosed less than 50 % of the time, while valve clearance issues were more reliably detected.

Previous work has laid a strong foundation and shown great potential for using audio signals as a vehicle diagnostic technique, with the capacity for smartphones to serve as capable diagnostic tools within the reach of the general public. Indeed, in past studies we have utilized internal smartphone sensors for a variety of automotive applications, from wheel imbalance detection [22] to tire pressure monitoring [21]. Thanks to mobile computing and pervasive sensing, there is an opportunity to help vehicle owners passively supervise the operation and

maintenance of their cars without environmental control or specialized equipment, yielding accuracy meeting or exceeding that of a trained and certified mechanic.

3   Data Collection

3.1   Experimental Design

The goal of the experiment was to collect audio in a manner that could reasonably be duplicated by typical vehicle owners with access to a smartphone. To that end, the procedure did not rely on fixed position or orientation of the vehicle or mobile phone, and the background environment was not controlled, which allowed ambient sources such as wind and other vehicles to add noise to both the training and testing data. In effect, we applied non-invasive and uncontrolled data collection.

To record the audio samples, each vehicle was warmed up for at least five minutes to ensure the engine was no longer running a "fast idle," which could provide unwanted audio artifacts. Then, the vehicle's hood was opened and propped up. Opening the hood allowed clearer audio signal capture during the proof-of-concept phase, and is something most drivers can easily complete without guidance or the use of tools.

For between two minutes and thirty seconds and six minutes, we used a mobile phone to record the engine idle sound as an uncompressed stereo. WAV file at 48000 Hz. During this time, mobile device was swept over the engine to provide a robust training set that incorporated noise from the engine intake, exhaust, belts, and other periodic signals present in the engine compartment. This relative motion is shown in Fig. 1.

With baseline testing completed, the procedure was repeated for anomalous engine operation and misfires. To simulate a misfire, the engine coil pack was

Fig. 1. The phone recorded as it was moved over the engine to provide background noise to test algorithm robustness. Engine covers were left on to minimize prep work and provide a better representative use case for in-situ monitoring.

Fig. 2. The supply to the ignition coil pack was disconnected in order to induce a complete misfire on individual cylinders.

disconnected with the engine turned off, removing the 12 V supply. This connector is shown for two vehicles in Fig. 2. Misfires induced in this manner manifest identically to misfires caused by coil failure, broken spark plug wires, fouled spark plugs, and improper grounding.

The engine was allowed to run for two minutes in this configuration prior to recording in order to allow the engine time to adapt to a cycle with periodic non-ignition. We selected two minutes as a lower limit because many engine control parameters, such as "long term fuel trim," reference only 30 seconds of driving history. In all cars, at least one cylinder was "deactivated" via induced misfire; in some cars, data were collected for multiple cylinders misfiring individually and in aggregate. After testing, the engine was shut off and the coil pack reconnected. If a check engine light had illuminated during testing, it was cleared using a standard ELM327-based automotive diagnostic tool.

Audio data were collected from multiple vehicles with different engine configurations over several days and in different parking locations (outdoor parking lots, a garage, and indoor parking structures). This allowed for the creation of a rich training set capable of providing in-data and out-data for testing. In the case of this experiment, the two engine configurations tested were a normally aspirated inline-four cylinder layout in a Kia Optima and a Ford Focus, as well as a normally aspirated V6 configuration in a Chevrolet Traverse and Nissan Frontier SUV. In cases where the engine cover had been removed to disconnect the coil pack, the cover was replaced prior to recording to better replicate a typical misfire condition wherein the engine's exterior features remain unperturbed.

4   Audio Analysis and Engine State Classification

Armed with audio samples from vehicle engines, we employed several data mining techniques in an attempt to detect and classify misfire occurrences. The

detection task was formulated as a supervised learning problem, and to simplify initial algorithm development, the audio samples were classified over only two operational states (normal and anomalous) as opposed to three or more (normal and different cylinder misfire configurations).

4.1   Feature Construction

The 48 kHz audio samples were first assigned labels based on whether the engine was operating normally or abnormally (an engine operating with a single misfiring cylinder) during recording. These samples were merged from stereo into a single, mono channel via averaging and the averaged samples were then subdivided into 2.5s segments. The first 1s and the last 2s of each audio sample were discarded to reduce noisy edge effects and clips with poor signal strength caused as a result of manipulating the mobile device.

2.5s samples recorded at 48 kHz correspond to 120, 000 discrete signal elements. The total number of samples in our data set was 992, out of which 373 corresponded to a normal engine. Figure 3 shows a segment of a normal engine audio signal along with that of a misfiring engine.
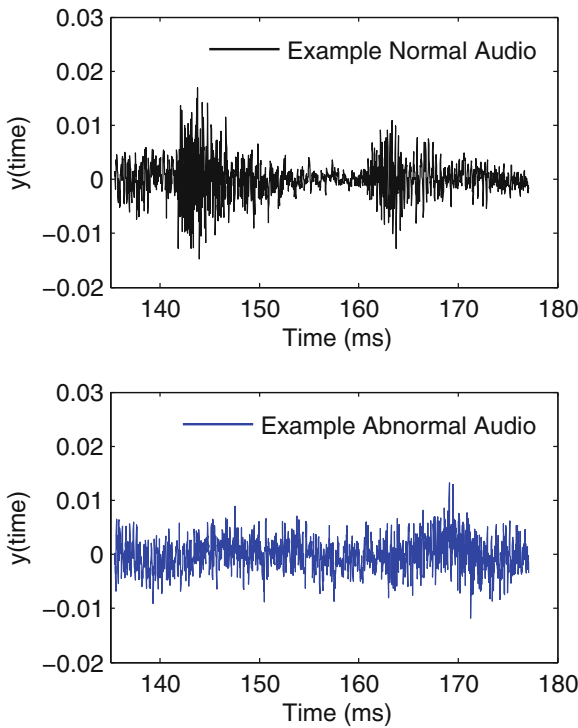

Fig. 3. Comparison of a segment of a normal audio sample with a misfiring audio sample.

To generate features for use in classification, each 120, 000 discrete signal element was then converted into a feature vector. We sought to generate a range of features to allow classification without the need for targeted, hypothesis-driven feature creation. Three classes of feature construction were employed and concatenated to form a long feature vector. The three classes include binned Fourier Transform coefficients, Wavelet Transform coefficients, and Mel Frequency Cepstral coefficients.

Though dense feature generation is an intensive process, this approach was adopted to remove any preconceived bias on what features have good discriminative power and rather allow machine learning techniques to drive the solution towards a reduced size feature set. A reasonable feature set size will allow rapid computation using the programmable Digital Signal Processors (DSPs) on mobile devices. Use of such processors has been shown to minimize a classification algorithm's impact on battery life significantly, even allowing Cloud-free operation [11], though further studies are required to optimize DSP computation and data transmission to the cloud in order to minimize overall power consumption and enable pervasive sensing with minimal annoyance to drivers.

Binned Fourier Transform (FT) Coefficients. The discrete samples were first normalized based on power and detrended to remove bias and linear drift. The Fast Fourier Transform (FFT) was then applied to convert the detrended time-domain signals into the frequency domain. Frequencies $< 10\,\text{kHz}$ were divided into bins 10 Hz wide. Higher frequencies were discarded as not providing additional differentiation because on average they comprised $< 25\,\%$ of the total energy and typically represented harmonics of lower frequencies. The average FT magnitude in each bin provided one feature. This process resulted in the creation of a feature vector of size 1000.
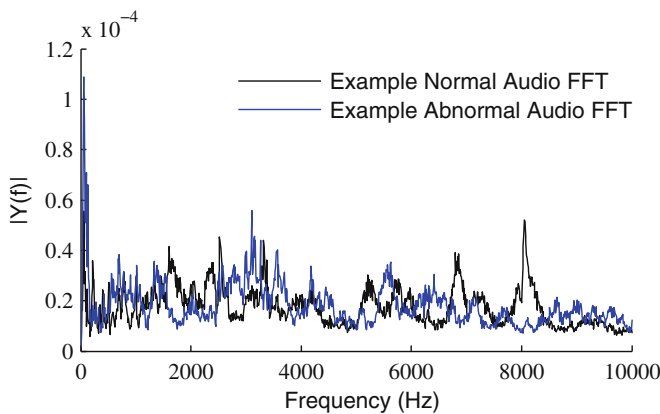

Fig. 4. Comparison of the spectral density of a normal audio sample with a misfiring audio sample.

Figure 4 shows an example comparison of the magnitude of the FT of a normal engine audio signal with a misfiring engine audio sample. We observe that several frequencies (in this particular example segment, around 2 kHz and 8 kHz) have a distinct pattern in the normal vs. abnormal cases. These frequencies that are statistically more powerful classifiers will be identified and used to classify a normal engine from a misfiring engine.

Discrete Wavelet Transform (DWT) Coefficients. In addition to the binned FT, we conducted a wavelet decomposition at level 10 on the power normalized, detrended discrete signal using Daubechies 4 wavelet. At each level of signal decomposition, mean, standard deviation and skewness was computed resulting in a 33-dimensional feature vector.

Mel Frequency Cepstral Coefficient (MFCC). The MFCC creates a spectral signature of short-term frames of the original signal that has been successfully applied to speech recognition [13]. We used a frame size of 1024 samples, with each frame incrementally shifted by 512 samples leading to a total number of 233 frames. For each frame 12 MFCC coefficients were extracted to form a feature vector of size 2796. We made use of the GNU-licensed Voicebox toolbox for MATLAB to conduct MFCC feature extraction.[1]

Concatenating the three sets of feature vectors from the FT, DWT, and MFCC resulted in a 3829-dimensional feature representation of the audio signal and a data matrix of size 992×3829. The data set was randomly divided into a 50 % training set and a 50 % test set. In most cases, samples of each state were drawn from different recording events. Rarely, segments of the same file may have been used in both training and testing. In such cases, the movement of the mobile device minimized the likelihood that samples were taken from similar locations and orientations, reducing sample dependence. After splitting the segments, subsequent work continued to develop appropriate feature reduction and classification techniques.

4.2   Feature Selection

To simplify computation, reduce redundancy and training time, and minimize overfitting, it was necessary to reduce the higher-dimensional feature vector using feature selection techniques [6,10,12]. Two filter-based methods were used for feature ranking: Fisher Score (FS) and Relief Score (RS) [10]. The use of feature ranking methods provides novelty over the state-of-the-art in audio classification for automotive faults, and will become instrumental in enabling low-power and resource-constrained devices to run this type of classification by eliminating the need to generate certain features.

---

[1] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.

The Fisher score [10] of a feature for binary classification is calculated using:

$$FS(f_i) = \frac{n_1(\mu_1^i - \mu^i)^2 + n_2(\mu_2^i - \mu^i)^2}{n_1\left(\sigma_1^i\right)^2 + n_2\left(\sigma_2^i\right)^2} \qquad (1)$$

$$= \frac{1}{n}\frac{(\mu_1^i - \mu_2^i)^2}{\frac{\left(\sigma_1^i\right)^2}{n_1} + \frac{\left(\sigma_2^i\right)^2}{n_2}},$$

where n_j is the number of samples belonging to class j, $n = n_1 + n_2$, μ^i is the mean of the feature f^i, μ_j^i and σ_j^i are the mean and the standard deviation of f_i in class j. A larger value corresponds to a variable having higher discriminating power.

The Relief score [12] of a feature is computed by first randomly sampling m instances from the data and then using:

$$RS(f_i) = \frac{1}{2}\sum_{k=1}^{m} d\left(f_k^i - f_{NM(\mathbf{x}_k)}^i\right) - d\left(f_k^i - f_{NH(\mathbf{x}_k)}^i\right), \qquad (2)$$

where f_k^i denotes the value of the feature f_i on the sample $\mathbf{x}_k$, $f_{NH(\mathbf{x}_k)}^i$ and $f_{NM(\mathbf{x}_k)}^i$ denote the values of the nearest points to $\mathbf{x}_k$ on the feature f_i with the same and different class label respectively, and $d(.)$ is a distance measure which was chosen to be the ℓ_2 norm. Here again, a larger score indicates a higher discriminating power of the variable.

Figure 5 shows the normalized score (scaled to $\in [0, 1]$) computed by the two methods noted above for each of the generated features. Though there is a significant correlation between the weights of FS and RS (a linear correlation coefficient of 0.49), combining the information from the two methods may reduce the likelihood of overfitting. To achieve this, we take a simple average of the scores from the two methods, calculated by:

$$AS(f_i) = \frac{1}{2}\left(\frac{FS(f_i) - \min(FS(f_i))}{\max(FS(f_i)) - \min(FS(f_i))} + \frac{RS(f_i) - \min(RS(f_i))}{\max(RS(f_i)) - \min(RS(f_i))}\right). \qquad (3)$$

With the features scored, we performed a systematic feature reduction study in order to identify a suitable subset of features. These feature subsets were parametrized by a variable p, with all features whose scores were in the top $(100 - p)^{th}$ percentile for discrimination were included in the subset. Figure 6 demonstrates how feature weighting varied with the FS, RS, and AS methods.

Figure 7 shows the variation in the 10−fold Misclassification Error Rate (MCR) on the training set using a linear Support Vector Machine (SVM), as well as the feature set size (#F) for different scoring schemes and the percentile cutoff p. We performed a grid search to find the optimal box constraint hyperparameter (C) for each of the feature subsets in the figure. From inspection, we identified a minimum MCR at $p = 90$ for each of the three feature scoring methods. Selection of a lower p results in a higher number of less informative features
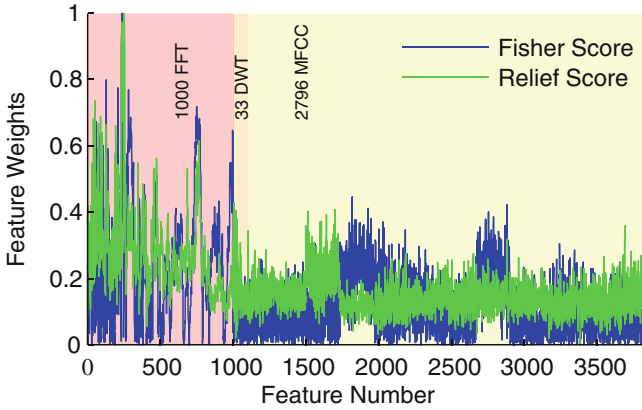
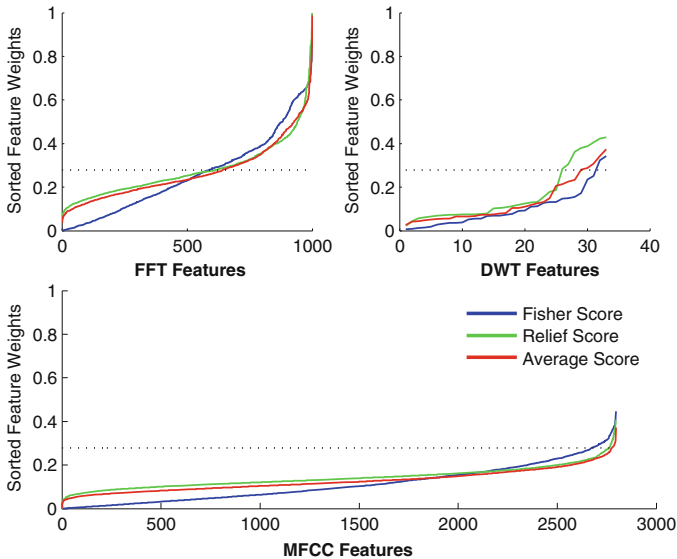**Fig. 5.** Comparison of the feature score calculated by the Fisher and the Relief Score methodology.

Fig. 6. Feature selection illustrated by method and type. For the AS method, $p = 90$ selection cutoff threshold is indicated as dotted black line at $w = 0.2784$.

in the subset, leading to overfitting and poorer cross-validation performance. Use of a higher p removes important features from the subset leading to a weaker model with decreased accuracy. We additionally observe that with the AS feature ranking the MCR increases less sharply after $p = 90$ when compared to FS or RS, likely due to variance reduction by model averaging. For these reasons, we selected AS with $p = 90$ as the optimal feature subset selection criterion.
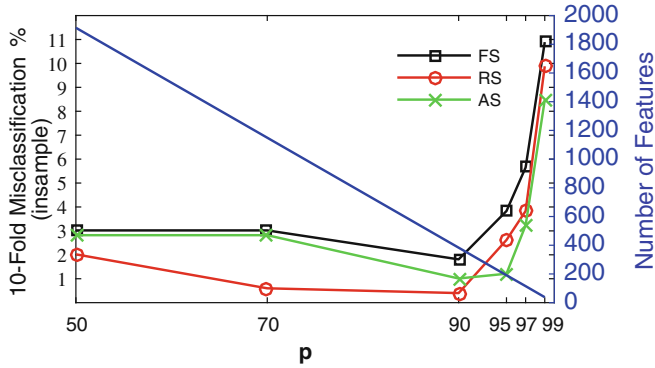
Fig. 7. Comparison of the 10−fold misclassification rate (MCR) and the feature set size with the variation in p.

The binned FT features alone result in a 10−fold misclassification rate of 0.8 %, while with the DWT the error is 36 % and the MFCC based features provide a 29 % error. Concatenating all the above features results in a misclassification rate of 2.6 % which is higher than FT alone. The minimum misclassification rate with FS, RS and AS scoring is 1.8 %, 0.4 % and 1.0 % respectively (Fig. 7).

The FT features have a higher discriminating power when compared to the other two classes of features. Simply combining the features from all three does not provide more discrimination than using the FT features. The ability to perform feature ranking and selecting the optimal subset improves the ratio of the discriminating power to the feature set size (i.e. (1-MCR)/#F) and therefore helps determine a small feature set with high discriminative power. It is also noted that the feature subset with AS and $p = 90$ has 358 FT features out of a total of 383 features, 5 DWT features and 20 MFCC features. Among the FT features selected from aggregate data, several were found to group around the 2.5 kHz and 7.5 kHz frequency bands.

4.3   Classification Algorithms

Using the chosen reduced feature set (AS feature weighting with $p = 90$ and $100 − p =$ the top 10th percentile of features selected), several classification algorithms were studied. The hyperparameters of the classification algorithms were optimized by conducting a grid search to minimize 10−fold cross-validation on the training data. The algorithms tested were k-Nearest Neighbor, Adaboost and SVM with linear, quadratic and RBF kernels. We found that for the SVM with the quadratic kernel all choices of the hyper parameter box-constraint cost (C) led to the same 10-fold misclassification error while for the RBF kernel the error sharply dropped from 38 % to 0 % around the optimal grid points (for finding C and γ). We therefore decided to remove the quadratic and RBF SVM from the final list of classifiers because we were unable to find a robust set of optimal hyperparameters.

5   Results and Conclusions

Table 1 summarizes the performance of the different classification algorithms on the 50 % outsample data. We observe that the linear SVM significantly outperforms the knn and Adabosst classification algorithms. With linear SVM, we obtained a misclassification rate of 1.0 % and the confusion matrix shown in Table 2. The 99 % accuracy of our approach well exceeds the prior art, indicating that our feature selection and reduction techniques are effective at not only improving algorithm efficiency, but increasing accuracy as well.

Considering that the reduced feature set is primarily comprised of the FT features, we trained a linear SVM (with C = 0.01) using only the FT features contained in the final reduced set from the previous section. The outsample misclassification rate with the top FT features was a slightly higher 1.2 % when compared to the results with using the top features of all types (see Table 1). This indicates that most of the discriminative information is contained in the FT features, with the DWT and MFCC features helping primarily differentiate edge cases. This presents an interesting trade off between computing cost and accuracy which will be relevant for designing a mobile application employing this technique. Current efficient implementations of FFT on smartphones [3] can be directly implemented for constructing the FT features in our reduced feature set, while there exist fewer algorithms to efficiently generate DWT and MFCC features.

Table 1. This table compares the classification accuracy (misclassification rate, reported-normal-when-abnormal false positive rate) for different tested algorithms.

Classifier Type	Optimal Hyperparameters	Misclassification Rate	False Positive Rate (Abnormal as Normal)
kNN (l_2 distance based)	Number of neighbors (n) = 11	25 %	27 %
Adaboost (learning rate = 0.3)	Tree depth = 7, Number of trees = 70	15 %	11 %
Linear SVM	Box-constraint cost (C) = 0.01	1.0 %	1.6 %

Table 2. The confusion matrix shows promising results for misfire detection, with 1.6 % false positives (reported normal when actually abnormal). We achieve similarly strong performance for false negatives (reporting abnormal when actually normal), potentially saving drivers money on unnecessary repairs.

		Predicted	
		Normal	**Abnormal**
Actual	**Normal**	100.0%	0.0%
	Abnormal	1.6%	98.4%

Finally, we note that in only one of the four vehicles did a "check engine" light come on at any point during testing, indicating that audio detection such as the one presented here with high accuracy and sensitivity may lend itself to the identification of a misfire prior to detection by an On-Board Diagnostic system. Early detection facilitates proactive response, and can help to lower vehicle maintenance and operating costs relative to drivers relying on the reactive diagnostic systems found in cars today.

6   Future Work

As a component of future work, we intend to explore the resource savings (computational and power) afforded by working with a reduced feature set. We have shown that feature ranking techniques facilitate the discarding of features with minimal loss in accuracy. These unused features need not be computed, enabling more efficient implementations of our feature generation algorithms suited to the limited resources found on mobile devices. Additionally, improving the off-line efficiency of these algorithms will allow us to develop an improved on-line approach, by minimizing bandwidth used for unnecessary data transmission and decreasing reference database size.

While this paper demonstrates promising results for the use of a mobile phone as a pervasive automotive diagnostic tool, the classification can be enriched and robustness improved to yield a more beneficial application, namely identification of the misfiring cylinder itself. That was difficult to discern in this study, as we suspect that information to be embedded within phase-based audio features, which are difficult to discern without a reliable indexing feature in the audio relative to engine component rotations. Other, non-combustion sounds are as of yet ill-defined (considering amplitude/frequency spread) and not available as a phase reference. Similarly, with the collected data it was not immediately feasible to distinguish among various anomalous misfire configurations, but we aim to study other techniques which may be used to improve differentiation among failed states. Such approaches may also improve classification of faults with lesser-defined signals, such as partial misfires due to lean conditions, and non-misfire faults such as clogged air filters or exhaust leaks.

To account for background noise, we intend to build a model to determine dependency of the audio waveform on the engine configuration (idle speed, cylinder count, aspiration, displacement, and firing order). Additionally, audio samples will be recorded from within the car to test whether the application can function from inside the vehicle.

Providing further data to enrich classification, the authors intend to develop algorithms for differential diagnosis: for example, measuring the sound near the air intake and exhaust to monitor airflow issues, identifying where in the airflow process an issue might be occurring. Finally, integrating audio data with information from the On-Board Diagnostic system may be possible, yielding richer fault information than is possible with either system alone.

References

1. Anami, B.S., Pagi, V.B.: Multi-stage acoustic fault diagnosis of motorcycles using wavelet packet energy distribution and ann. In: SERSC International Journal of Advanced Science and Technology (December 2012), International Journal of Advanced. Science and Technology **49**, 47–62 (2012)
2. CarMD: 2011 CarMD Vehicle Health Index. https://www.carmd.com/wp/vehicle-health-index-introduction/2011-carmd-vehicle-health-index/
3. de Carvalho Jr., A.D., Rosan, M., Bianchi, A., Queiroz, M.: Fft benchmark on android devices: Java versus jni. Nexus **7**, 1 (2013). http://compmus.ime.usp.br/sbcm/2013/pt/docs/pos_tec_4.pdf
4. Dandare, S.N.: Multiple fault detection in typical automobile engines: a soft computing approach. WSEAS Trans. Signal Process. **9**(3), 158–166 (2013)
5. Galloni, E.: Dynamic knock detection and quantification in a spark ignition engine by means of a pressure based method. Eng. Convers. Manag. **64**, 256–262 (2012)
6. Gheyas, I.A., Smith, L.S.: Feature subset selection in large dimensionality domains. Pattern Recogn. **43**(1), 5–13 (2010)
7. IHS Inc: Aging vehicle fleet continues to create new opportunity for automotive aftermarket, ihs says. http://press.ihs.com/press-release/automotive/aging-vehicle-fleet-continues-create-new-opportunity-automotive-aftermarket
8. Kabiri, P., Ghaderi, H.: Automobile independent fault detection based on acoustic emission using wavelet. In: Singapore International NDT Conference and Exposition 2011, Singapore International NDT Conference and Exposition, Singapore, November 2011
9. Kabiri, P., Makinejad, A.: Using PCA in acoustic emission condition monitoring to detect faults in an automobile engine. In: 29th European Conference on Acoustic Emission Testing (EWGAE2010), pp. 8–10 (2011)
10. Kumar, S.: Mobile sensor systems for field estimation and "hot spot" identification. Ph.D. thesis, Massachusetts Institute of Technology (2014)
11. Lane, N.D., Georgiev, P., Qendro, L.: Deepear: robust smartphone audio sensing in unconstrained acoustic environments using deep learning. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 283–294. ACM (2015)
12. Liu, H., Motoda, H.: Feature selection for knowledge discovery and data mining, vol. 454. Springer Science & Business Media, Heidelberg (2012)
13. Logan, B., et al.: Mel frequency cepstral coefficients for music modeling. In: ISMIR (2000)
14. Merkisz, J., Bogus, P., Grzeszczyk, R.: Overview of engine misfire detection methods used in on board diagnostics. J. Kones Combust. Engines **8**(1–2), 326–341 (2001)
15. Merola, S.S., Vaglieco, B.M.: Knock investigation by flame and radical species detection in spark ignition engine for different fuels. Eng. Convers. Manag. **48**(11), 2897–2910 (2007)
16. Navea, R.F., Sybingco, E.: Design and implementation of an acoustic-based car engine fault diagnostic system in the android platform. In: International Research Conference in Higher Education. Polytechnic University of the Philippines (2013)
17. Regulation, section 1968.2 malfunction and diagnostic system requirements - 2004 and subsequent model year passenger cars
18. Schrank, D., Eisele, B., Lomax, T., Bak, J.: 2015 urban mobility scorecard. Texas A & M Transportation Institue & INRIX (2015). http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-scorecard-2015.pdf

19. Service, B.U: Tech feature: Detecting misfires in OBD II engines. http://www.underhoodservice.com/tech-feature-detecting-misfires-in-obd-ii-engines/
20. Siegel, J.E.: Data Proxies, the Cognitive Layer, and Application Locality: Enablers of Cloud-Connected Vehicles and Next-Generation Internet of Things. Ph.D. thesis, Massachusetts Institute of Technology (2016)
21. Siegel, J.E., Bhattacharyya, R., Desphande, A., Sarma, S.E.: Smartphone-based vehicular tire pressure and condition monitoring. In: Proceedings of SAI Intellisys 2016 (2016)
22. Siegel, J.E., Bhattacharyya, R., Sarma, S., Deshpande, A.: Smartphone-based wheel imbalance detection. In: ASME 2015 Dynamic Systems and Control Conference. American Society of Mechanical Engineers (2015)
23. Sujono, A.: Utilization of microphone sensors and an active filter for the detection and identification of detonation (knock) in a petrol engine. Mod. Appl. Sci. **8**(6), 112 (2014)
24. Tse, P.W., Tse, Y.L.: On-road mobile phone based automobile safety system with emphasis on engine health evaluation and expert advice. In: Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET 2012, pp. 3232–3241. IEEE (2012)
25. United States Department of Transportation, Federal Highway Administration: traffic volume trends. https://www.fhwa.dot.gov/policyinformation/travel_monitoring/15dectvt/
26. Vulli, S., Dunne, J.F., Potenza, R., Richardson, D., King, P.: Time-frequency analysis of single-point engine-block vibration measurements for multiple excitation-event identification. J. Sound Vibr. **321**(3), 1129–1143 (2009)
27. Zhang, Z., Saiki, N., Toda, H., Imamura, T., Miyake, T.: Detection of knocking by wavelet transform using ion current. In: ICICIC, pp. 1566–1569. IEEE (2009)