Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 1807 – 1826

doi:10.3906/elk-1605-77

Automobile engine condition monitoring using sound emission

Hamid GHADERI, Peyman KABIRI*
Intelligent Automation Laboratory, School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran

Received: 07.05.2016 • **Accepted/Published Online:** 25.06.2016 • **Final Version:** 29.05.2017

Abstract: A wavelet packet transform (WPT) is a well-known technique used for data and signal-processing that has proven to be successful in condition monitoring and fault diagnosis. In this study, using feature extraction based on wavelet transformation, sound signals emitted from automobile engines under both faulty and healthy conditions are analyzed. The intention is to categorize sound signals into both healthy and faulty classes. Sound signals are generated from 4 different automobile engines in both healthy and faulty conditions. The investigated fault is within the ignition system of the engines. In addition, there are other possible problems that may also affect the generated sound signals. In the reported study, a set of features is initially extracted from the recorded signals. The more informative features are later selected using a correlation-based feature selection (CFS) algorithm. Results prove the efficiency of wavelet-based feature extraction for the case study of the reported work.

Key words: Wavelet packet transform, condition monitoring, fault diagnosis, correlation-based feature selection

1. Introduction

The growth of the automobile industry requires cost-effective fault diagnosis and maintenance of automobiles. Consequently, fault diagnosis for internal combustion engines is an important issue. Therefore, the development of an automobile engine condition monitoring system capable of providing early warning about the engine's state of operation is needed. Both the vibration [1,2] and sound signals [3,4] have been widely used for fault diagnosis. Xian et al. [2] used a wavelet packet transform (WPT) for feature extraction in their reported work; however, the authors were targeting a vibration frequency band less affected by environment noise compared to sound. In their approach, no feature selection was used. Wu et al., in their reported work [4], focused on the sound of the engine, and a WPT was used to extract features for individual faults but no feature selection was reported. They used a rule-based system that is slower than our proposed system.

Figlus et al. [5] used the sound of the engine for automatic detection of enlarged clearance valves. In their work, the WPT was used for feature extraction. Once the WPT was applied maximum and minimum values were used for the detection. Only 2 sampled engines were used in the experiments. In another reported work [6], several faults were considered on only a single sample engine/automobile. There is also a reported work that focused on motorbike piston-bore fault identification [7].

Our study concentrates on single-fault detection in 4 different categories of car models using 60 different samples in each category (car model). In this way, the proposed approach is expected to be applicable for a wide range of car categories; hence, it is more general in its capabilities. Operating in hearing range and

*Correspondence: peyman.kabiri@iust.ac.ir

sound frequency also proves that our approach can handle environmental noise efficiently and can make use of inexpensive data collection equipment. Both sound and vibration signals generated by the engines contain considerable dynamic information about the engine's operating conditions [8]. The characteristics of the sound signal generated by a mechanical system can be represented by a typical sound waveform. The typical waveform associates with the state of operation for the monitored components within the device. Thus, signal-processing techniques can provide useful methods for fault diagnosis and condition monitoring. Extracting information or features that are closely related to a specific fault is a great challenge in fault diagnostic and condition monitoring based on sound signal-processing techniques. Furthermore, the nature of the signals and the accuracy that results from the extracted information determines the suitability of signal-processing techniques [9]. Many signal-processing techniques have been proposed in the literature. For example, the fast Fourier transform (FFT) is a frequency domain analysis that has been used to extract frequency domain features [10]. This method relies on variations in frequency to identify different faulty conditions. The FFT transfers the signals to the frequency domain with a loss of information occurring in the process. While implementing the FFT, signals are assumed to be stationary or, in other words, their statistics do not change over time. The FFT of a nonstationary signal is the frequency domain averaged over the duration of the signal. Consequently, the FFT cannot represent properties of transient signals at lower frequencies. Moreover, the FFT suffers from vulnerability to background noise [11]. The short-time Fourier transform (STFT) was introduced as an alternative to the FFT in order to overcome its aforementioned drawbacks. The STFT uses a window function to divide the signal into small segments and then calculates the FFT of each segment to evaluate its frequency and phase content. Since the width of the window function is fixed for the entire duration of the signal, the STFT has a fixed time-frequency resolution. Recently, the wavelet transform (WT) has attracted many researchers due to its promising solutions for signal processing in time-frequency domains. The continuous wavelet transform (CWT) is used in the field of fault diagnosis [12,13]. The discrete wavelet transform (DWT) is noteworthy since it takes longer to calculate the CWT coefficients related to different resolutions of a signal [14,15]. The main advantage of the WT is that it uses a resizable window that becomes dilated for lower frequencies and sharpened for higher frequencies. This is due to the fact that higher frequencies of a signal require more detailed analysis in the frequency domain while lower frequencies of the signal require more detailed analysis in the time domain. Nevertheless, the conventional WT lacks precise analysis of the higher frequencies of signals. The WPT is a generalization of the traditional WT; that is, unlike the DWT, which only decomposes the approximation version of the signals, the WPT decomposes both approximation and detail versions. The WPT provides more valuable information about the signal.

Constructed datasets might contain irrelevant features, and irrelevant features may increase the complexity of the classification process and reduce classification accuracy. Hence, adopting a feature-reduction technique seems essential. Various feature-reduction techniques have been used by many researchers. For instance, independent component analysis (ICA) [16,17], the decision tree (DT) [14,18], the self-organizing map (SOM) [19], principal component analysis (PCA) [20], and the genetic algorithm (GA) with artificial neural network (ANN) [21–23] are a few of the feature-reduction techniques that have been used in the field of fault diagnosis.

In this study, it was decided to distinguish the faulty engine from a healthy one using sound signals generated from the engines under both healthy and faulty conditions. The investigated fault is within the ignition system of the engines, meaning that the engines are operating with one misfiring cylinder. Apart from the investigated fault, the engines might also suffer from other possible problems that may affect the generated sound signals. For example, one of the most common faults besides the investigated fault is the combustion-

timing process fault. There are 2 reasons for using a WPT for feature extraction. The first reason is to have a good signature for accurate representation of the sound signals generated under either of the 2 engine conditions. The second reason is because sound signals emitted from automobile engines are nonstationary [24]. Attaining the highest capability for the extracted features is basically dependent on the selection of the wavelet bases. Different alternatives are proposed to find the most appropriate wavelet bases. For example, a GA is used to find the scale parameter and some wavelet family-related parameters [25]. Nevertheless, the GA's outcome is not necessarily the best wavelet base. Another heuristic approach is to define the wavelet bases with respect to their similarity with the desired signals. In this approach, a large number of wavelet bases are investigated, and the classification results obtained by using them with a WPT are then reported. Using correlation-based feature selection (CFS) with the best-first algorithm (BF) as a feature subset search strategy, irrelevant features are ignored, and the most informative features are selected. The reduced datasets are used for classification. Classification results are validated and reported using 10-fold cross validation, where 10% of randomly selected samples are used for training and 90% for testing. Results show the efficiency of the wavelet-based feature extraction for the case study in this paper.

2. Materials and methods

2.1. Wavelet transform

A wavelet transform (WT) is used to analyze signals at different frequency resolutions. The WT does this using its multiresolution analysis (MRA) framework. The MRA provides good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. To do so, the WT uses a window function with various widths. The window function is either called a mother wavelet or a wavelet base. This function is a small wave with a finite-length (compactly supported) oscillatory and acts like a prototype for generating other window functions that are used in the transform with different regions of support. Known as time-frequency or time-scale analysis, the WT is a function of 2 independent variables called scale and translation. The translation parameter refers to the location of the mother wavelet as it is shifted through the entire duration of the signal. This parameter corresponds to the time information of the signal. The scale parameter refers to the stretch (or compression) of the mother wavelet. Low scales correspond to compressed mother wavelets and vice versa. In the frequency domain, the compressed mother wavelet better shows the high frequency information of the signal, and the dilated mother wavelet shows the low frequency information of the signal, as well. The WT is divided into a continuous and a discrete transform. In the continuous version, i.e. the CWT, the 2 aforementioned parameters are continuously changing. However, in the discrete version, i.e. the DWT, a subset of every possible scale and translation values is selected. In other words, the dyadic scales and translations are selected, which means that the scales and translations are in the power of 2. This makes the discrete version resemble a filter bank through which the original signal is passed. Eq. (1) is a representation of any orthogonal function f. Eq. (2) is yet another form representing the original signal construction.

$$f = \sum_{x} \langle f,\ x \rangle\ x \tag{1}$$

$$\begin{aligned} y &= \sum_{k} \alpha_k x_k \\ \alpha_k &= \langle x_k\ ,\ y \rangle \end{aligned}\ , \tag{2}$$

where $\langle f,\ x\rangle$ is the inner product of function f and of basis function x. In Eq. (1) or Eq. (2), function f can be represented by a wavelet basis function such as Eq. (3).

$$\begin{aligned} f(t) &= \sum_{J,\,k} \left\langle f(t),\ \psi\left(2^J t\ -\ k\right)\right\rangle \psi\left(2^J t\ -\ k\right) \\ &= b_0\varphi\left(t\right)\ +\ \sum_{J\ ,\ k} b_{2^J+k}\ .\ \psi\left(2^J t\ -\ k\right) \end{aligned}\quad , \tag{3}$$

where ψ are the dilated and translated basis functions, and ϕ is the scaling function.

A DWT is performed using high-pass and low-pass filters. The high-pass filter generates the high frequency (low-scale) information of the signal called detailed coefficients, and the low-pass filter generates the low frequency (high-scale) information of the signal called approximation coefficients. The DWT decomposes the signal into different decomposition levels. The decomposition level represents the number of times the original signal passes through the transforming filters. Each new decomposition level is calculated by passing only the approximation coefficients computed from the previous decomposition level. The detailed coefficients are left out. To involve the detailed coefficients in the transform, a WPT is used as described in the following section.

2.1.1. Wavelet packet transform (WPT)

Recently, a WPT was used by researchers in areas of fault diagnosis such as automotive engine ignition systems [26,27], rolling element bearings [28,29] and gearboxes [30]. A WPT is a generalization of a conventional DWT that provides more detailed signal-processing capabilities. Both WPTs and DWTs contain multiresolution analysis (MRA) frameworks. Unlike the DWT, which only decomposes the approximation version of the signals, the WPT decomposes both approximation and detail versions. Therefore, the WPT has the same frequency bandwidths in each resolution while the DWT does not have this characteristic. The WPT iteratively decomposes the signal via 2 digital filters, and then the outputs are down-sampled by a factor of 2. The filters used are low-pass and high-pass filters known as scaling function and discrete mother wavelet, respectively. The low-pass filter extracts the high-frequency components, and the high-pass filter extracts the low-frequency components of the original signal. Both mother wavelet and scaling function must be determined before the transformation is performed. Wavelet packet function can be described using Eq. (4) [4]:

$$w^n_{j,\,k}(x)\ =\ 2^{j\,/\,2} w(2^j x\ -\ k),\quad j,\ k\ \in\ \mathrm{Z}, \tag{4}$$

where j and k are integers indicating the index scaling and translation operations. The index n denotes the oscillation parameter.

The first 2 wavelet-packet functions are mother wavelet and scaling function [4]:

$$w^0_{0,\,0}(x) = \varphi(x) \tag{5}$$

$$w^1_{0,\,0}(x) = \psi(x), \tag{6}$$

where $n=$ 2, 3, Eq. (4) can be rewritten as Eq. (7) and Eq. (8) [4]:

$$w^{2n}_{0,\,0}(x)\ =\ \sqrt{2}\sum_{k\in\mathrm{Z}} h(k)\ w^n_{1,\,k}(2x\ -\ k) \tag{7}$$

$$w_{0,\,0}^{2n\,+\,1}(x) \;=\; \sqrt{2}\sum_{k\in\,\mathrm{Z}} g(k)\; w_{1,\,k}^{n}(2x \;-\; k), \tag{8}$$

where $h(k)$ and $g(k)$ are low-pass and high-pass filters, respectively. These 2 filters are associated with the predefined scaling function and mother wavelet function, respectively. They are known as quadratic mirror filters (QMFs).

The wavelet-packet coefficients $w_{j,\,k}^{n}$ are calculated through the inner product $\left\langle f(x),\; w_{j,\,k}^{n}\right\rangle$ defined by Eq. (9) [4]:

$$w_{j,\,k}^{n} \;=\; \left\langle f(x),\; w_{j,\,k}^{n}\right\rangle \;=\; \int f(x)\; w_{j,\,k}^{n}(x)\; dx \tag{9}$$

The full wavelet-packet tree with 3-level decomposition in filter-bank ordering [31] is depicted in Figure 1.

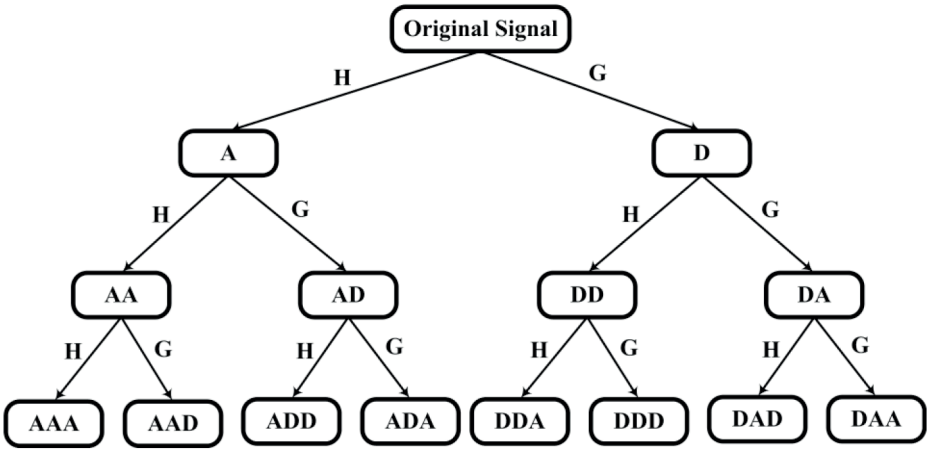

Figure 1. The 3-level WPT decomposition in filter-bank ordering.

The letters A and D stand for approximation and detail coefficients, respectively, and H and G stand for low-pass and high-pass filters, respectively.

The frequency range of each node is calculated using Eq. (10) [32].

$$\left((n \;-\; 1)\;\times\; 2^{-j\,-\,1}\;\times\; F_s,\; n\;\times\; 2^{-j\,-\,1}\;\times\; F_s\right],\quad n\;=\;1,\;2,\;...,\;2^{j}, \tag{10}$$

where F_s is the sampling frequency and j is the decomposition level. In the reported work, the sampling frequency used to record and analyze the engine sound signals is $F_s \;=\; 44100\;Hz$, which indicates that the covered frequency range is [0–22050] Hz. For a better understanding of the frequency ranges during the decomposition, filter-bank ordering [31] is used in Figure 1, *natural frequency ordering* [31] is used in Figure 2, and Table 1 shows the frequency range of each wavelet packet.

2.2. Feature extraction

In the reported work, sound signals are decomposed by a WPT into 3 levels. Therefore, the last level of the wavelet packet tree will have $2^3 \;=\; 8$ nodes. In order to find signatures for the analyzed sound signals, the

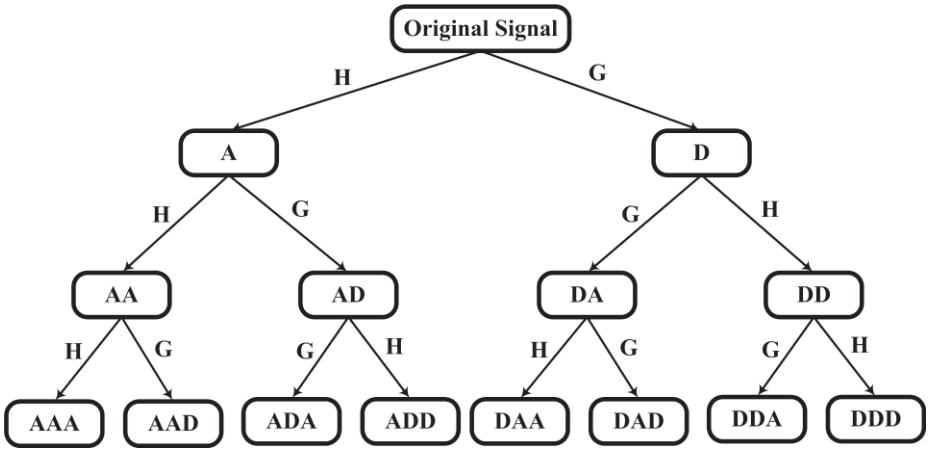

Figure 2. The 3-level WPT decomposition in natural frequency ordering.

Table 1. Frequency range of each wavelet packet for each decomposition level.

Wavelet Packet	Lower frequency (Hz)	Upper frequency (Hz)	Center frequency (Hz)
A	0	11025	5512.5
D	11025	22050	16537.5
AA	0	5512.5	2756.25
AD	5512.5	11025	8268.75
DA	11025	16537.5	13781.25
DD	16537.5	22050	19293.75
AAA	0	2756.25	1378.125
AAD	2756.25	5512.5	4134.375
ADA	5512.5	8268.75	6890.625
ADD	8268.75	11025	9646.875
DAA	11025	13781.25	12403.125
DAD	13781.25	16537.5	15159.375
DDA	16537.5	19293.75	17915.625
DDD	19293.75	22050	20671.875

following features are extracted from each wavelet packet decomposition tree node:

$$Energy_{j,\,n} \;=\; \sum_{k} \left| w^{n}_{j,\,k} \right|^{2} \tag{11}$$

$$Kurtosis_{j,\,n} \;=\; \frac{\sum\limits_{k} \left(w^{n}_{j,\,k} \;-\; \overline{w^{n}_{j,\,k}} \right)^{4}}{(K \;-\; 1)\sigma^{4}} \;-\; 3 \tag{12}$$

$$Max_{j,\,n} \;=\; \max_{k} \left(w^{n}_{j,\,k} \right) \tag{13}$$

$$Min_{j,\,n} \;=\; \min_{k} \left(w^{n}_{j,\,k} \right) \tag{14}$$

Calculating the aforementioned features from each node within the last level of decomposition tree, a dataset with 32 features is constructed. While the constructed dataset may contain irrelevant features that adversely

affect classification accuracy, finding the most informative features is of great importance. The most informative features in the reported work were found using a CFS algorithm. The reason why energy, kurtosis, minimum, and maximum features are selected is because of the better experimental results in signal decomposition using wavelet transformation. The energy feature provided good results with the highest frequency resolution. This algorithm is explained in detail in the next section.

2.3. Utilized wavelets

In the reported work, Daubechies, Symlet, and Coiflet wavelets were used. From a smoothness and noise cancellation point of view, the selection of the appropriate type of wavelet can improve results, considering that an increase in the order of the mother wavelet will result in extraction of more complex information. Experiments show that low-order mother wavelets produce more error in fault detection. Therefore, the intension was to increase the frequency resolution at higher frequencies. Increasing this order to a much higher order may not be helpful either. Experiments show that the type of mother wavelet is a more effective parameter than the order of the mother wavelet. In this work, only 3 types of wavelets were considered. This was due to the short time period available to do the work.

Using a dilation equation for a wavelet of order N, the mother wavelet can be produced by Eq. (15), and the scaling function is derived from Eq. (16).

$$\varphi(t) = \sum_{k=0}^{N-1} c_k \varphi(2t - k) \tag{15}$$

$$\psi(t) = \sum_{k=0}^{N-1} (-1)^k c_k \varphi(2t + k - N + 1), \tag{16}$$

where c_k are filter coefficients and both the $\varphi(t)$ and $\psi(t)$ functions should satisfy certain normalization and orthogonalization constraints.

2.3.1. Daubechies wavelets

These wavelets were first introduced by Ingrid Daubechies in 1988 [33]. The names of Daubechies family wavelets are constructed by substituting N in dbN, where N is the order of the wavelet. This family of wavelets does not have explicit expressions except db1 (or Haar). The support length of wavelet function and its corresponding scaling function is 2N-1. These wavelets are not symmetrical to N and have N vanishing moments for the mother wavelet (when the mother wavelet is equal to 0) but not for the scaling function (also called the father wavelet).

If vanishing moments (points) exist for the father wavelet, calculating the values of function f at discrete points, the wavelet coefficients are approximated with Eq. (17) [34]:

$$\alpha_{jk} = 2^{-j/2} f(\frac{k}{2^j}) + r_{jk}, \tag{17}$$

where r_{jk} is small enough.

The Daubechies mother wavelet function is presented in Eq. (18), and the scaling function is $\varphi(y)$, at which point its Fourier transform $\widehat{\phi}(y)$ satisfies Eq. (19) [35].

$$\psi(x) = \sum_k g_k \varphi(2x - k) \tag{18}$$

$$\widehat{\varphi}(y) = \left(1 \Big/ \sqrt{2\pi}\right) \prod_{k \geq 1} m_0\left(2^{-k} y\right), \tag{19}$$

where $m_0(y)$ is called a low-pass filter and a trigonometric polynomial defined by Eq. (20).

$$m_0(y) = \sqrt{2} \sum_k h_k e^{iky} \tag{20}$$

A finite sequence $h_0, h_1, \ldots, h_{2N-1}$ should be constructed to satisfy Eqs. (21)–(23).

$$\sum_k h_k h_{k+2n} = \delta_{n0} \quad for \quad \forall n \in \mathrm{Z} \tag{21}$$

$$\sum_k h_k = \sqrt{2} \tag{22}$$

$$\sum_k g_k k^n = 0, \tag{23}$$

whenever $0 \leq n \leq N-1$, where $g_k = (-1)^k h_{1-k}$. The Kronecker symbol δ_{nm} is equal to 1 for n = m and is 0 otherwise.

2.3.2. Symlet wavelets

These wavelets are called SymN, where N indicates the order. Symlets are modifications of Daubechies wavelets that are introduced to have more symmetry than Daubechies wavelets. Symlets are only near symmetric. The support length of the wavelet function and its corresponding scaling function is $2N-1$. These wavelets have N moments equal to 0.

No system of φ, ψ, other than the Haar system, can be compactly supported and at the same time also symmetric [36]. Nonetheless, from a practical point of view, one can try to make the system as close as possible to a symmetric system in the following way: minimizing the phase of $m_0(\xi)$ among all the $m_0(\xi)$ moments with the same $|m_0(\xi)|$ value [34]. Coefficients $\{h_k\}$ for the sampled symlets are presented in Table 2 [36].

2.3.3. Coiflet wavelets

Coiflets were originally designed by Ingrid Daubechies per Ronald Coifman's request. These new types of wavelet, i.e. coiflets, are reported on in a work by Beylkin et al. [37]. Coiflets have scaling functions (N/3-1) with vanishing moments (N/3) and are near symmetric. These wavelets are named CoifN, with N indicating the order. More symmetric than Daubechies wavelets, these wavelets have a support length equal to $6N--1$. Moreover, they have the highest number of moments equal to 0, i.e. 2N. Coiflets are orthonormal wavelet systems.

While constructing coifiets $m_0(\xi)$, Eq. (24) should be calculated as

$$m_0(\xi) = \left(\frac{1+e^{-i\xi}}{2}\right)^N \Im(\xi), \tag{24}$$

Table 2. Symlet coefficients for $N = 4$ and $N = 8$.

	N	$c_{N,n}$
N = 4	0	–0.107148901418
	1	–0.041910965125
	2	0.703739068656
	3	1.136658243408
	4	0.421234534204
	5	–0.140317624179
	6	–0.017824701442
	7	0.045570345896
N = 8	0	0.002672793393
	1	–0000428394300
	2	–0.021145686528
	3	0.005386388754
	4	0.069490465911
	5	–0.038493521263
	6	–0.073462508761
	7	0.515398670374
	8	1.099106630537
	9	0.680745347190
	10	–0.086653615406
	11	–0.202648655286
	12	0.010758611751
	13	0.044823623042
	14	–0.000766690896
	15	–0.004783458512

where $\Im(\xi)$ is a trigonometric polynomial. The following conditions should be satisfied:

$$\int \varphi(x)\, dx \;=\; 1, \int x^l \varphi(x)\, dx \;=\; 0\, l \;=\; 1,\; ...,\; N \;-\; 1 \tag{25}$$

$$\int \psi(x)\, x^l dx \;=\; 1\,,\; l \;=\; 0\,,\; ...\,,\; N \;-\; 1 \tag{26}$$

For a wavelet ψ with scaling function ϕ, Monzón et al. [38] define coiflets in the following way:

"If Let $\{h_j\}_{j\,=\,0}^{L\,-\,1}$ be the coefficients of a real QMF H. We say that H is a *coiflet* of shift α and moments M, N if the following three conditions are satisfied:

$$\sum_{j\,=\,0}^{L\,-\,1} (-\,1)^j j^k h_j \;=\; 0 \quad for\; 0 \;\leq\; k < \;M, \tag{27}$$

$$\sum_{j\,=\,0}^{L\,-\,1} j^k h_j \;=\; \alpha^k \quad for\; 0 \;\leq\; k \;<\; N, \tag{28}$$

$$3M \;>\; L \;-\; 1 \quad and\; 3N \;\geq\; L \;-\; 1\,. \tag{29}$$

They described the coiflet system in terms of a_k and b_k (Eq. (30)) [38]:

$$a_k = \frac{1}{k!}\sum_j \left(j - \frac{\alpha_0}{2}\right)^k h_{2j} \quad and \quad b_k = \frac{1}{k!}\sum_j \left(j - \frac{\alpha_0 - 1}{2}\right)^k h_{2j+1}, \tag{30}$$

where $0 \leq k \leq l$ and $l = \frac{1}{2}(L-2)$ [38]. For $k > l$, a_k and b_k can be expressed in terms of variables with $V = \{a_0, \ldots, a_l, b_0, \ldots, b_l\}$ [38].

2.4. Correlation-based feature selection (CFS)

Finding the most important set of features is considered to be a major issue in machine-learning algorithms. CFS partly overcomes the aforementioned problem using a correlation-based approach. A heuristic in a CFS algorithm is used to evaluate how worthy a subset of features is. This is called the merit of that feature subset. The heuristic is based on a hypothesis that a good subset of features consists of those that are uncorrelated with each other while being highly correlated with the class. The hypothesis is formulized in Eq. (27) [32]:

$$Merit_S = \frac{k\overline{r_{c,f}}}{\sqrt{k + k(k-1)\overline{r_{f,f}}}}, \tag{31}$$

where $Merit_S$ indicates the merit of the feature subset S containing k features. Parameters $\overline{r_{c,f}}$ and $\overline{r_{f,f}}$ indicate mean feature-class and mean feature-feature correlation, respectively, where $f \in S$. In Eq. (27), the numerator indicates how predictive the feature subset S is in predicting the class. The denominator indicates how much redundancy there is among the features in feature subset S. The more the features are correlated with each other, the more redundant they are. This heuristic discards irrelevant features that are poor class predictors.

It should be mentioned that the numerical features should be discretized before CFS is applied. Discretization is the process of transforming continuous-valued attributes to nominal ones. There are different discretization techniques by which the variation range of a specific feature is divided into a number of partitions. In the reported work, the discretization technique proposed by Fayyad et al. [39] was used. The applied discretization algorithm is based on a minimal entropy heuristic. This supervised algorithm uses class information entropy of candidate partitions to select the best partition boundary (known as cut point) for discretization. For a given set of instances such as S, a feature A, and a partition boundary T, the class information entropy of the partitions induced by T ($E(A, T; S)$) is calculated with Eq. (32) [39]:

$$E(A, T; S) = \frac{|S_1|}{|S|}Ent(S_1) + \frac{|S_2|}{|S|}Ent(S_2), \tag{32}$$

where S_1 and S_2 are 2 intervals of S bounded by T, and $Ent(S)$ is the class entropy of the set S calculated by Eq. (33) [39]:

$$Ent(S) = -\sum_{i=1}^{C} p(C_i, S)\log_2(p(C_i, S)), \tag{33}$$

where $p(C_i, S)$ is the probability of class C_i represented in set S.

The partition boundary that minimizes the class entropy function (Eq. (33)) over all of the possible partition boundaries is then selected. The selected partition boundary is called $T_{\min}$. This method can be

recursively applied to both partitions induced by $T_{\min}$ until a termination condition is met. The termination condition used by Fayyad et al. [39] is known as the Minimal Description Length Principle. This criterion indicates that a partition induced by a partition boundary T is accepted if and only if Eq. (34) [39] is true.

$$Gain(A,\ T;\ S)\ >\ \frac{\log_2(N\ -\ 1)}{N}\ +\ \frac{\Delta(A,\ T;\ S)}{N}, \tag{34}$$

where N is the number of instances in S and $Gain(A,\ T;\ S)$ and $\Delta(A,\ T;\ S)$ are calculated using Eqs. (35) and (36) [39], respectively.

$$Gain(A,\ T;\ S)\ =\ Ent(S)\ -\ E(A,\ T;\ S) \tag{35}$$

$$\Delta(A,\ T;\ S)\ =\ \log_2(3^k\ -\ 2)\ -\ [kEnt(S)\ -\ k_1Ent(S_1)\ -\ k_2Ent(S_2)]\,, \tag{36}$$

where, in Eq. (36), k_i is the number of class labels represented in set S_i.

After discretization, a set of nominal values is assigned to each partition. Correlation between the 2 features X and Y is measured by symmetric uncertainty (Eq. (37)) [40]:

$$Symmetric\ uncertainty\ =\ \frac{2\ \times\ gain}{H(Y)\ +\ H(X)}, \tag{37}$$

where $H(.)$ is the entropy of the feature and the nominator is the information gain that is a symmetric measure that equals the amount of information gained about Y after observing X. Entropy is considered to be a measure of uncertainty or unpredictability in a system. The entropy of feature X is calculated with Eq. (38).

$$H(X)\ =\ -\sum_{x\in X} p(x)\log_2(p(x)), \tag{38}$$

where $p(x)$ is the probability of the nominal values for feature X $(x \in X)$. The information gain or mutual information is calculated using Eq. (39) [41]:

$$gain\ =\ H(Y)\ +\ H(X)\ -\ H(X,\ Y) \tag{39}$$

If n possible features exist in the initial dataset, there can be 2^n possible subsets of features. There is only one way to find the best subset and this is to evaluate all of them. This approach is clearly not feasible for high dimensional datasets. Thus, a certain search strategy is needed to explore all the possible feature subset space. In this paper, a BF search strategy is used. Starting from an empty subset of features, a BF search tries to explore the feature space by making the local changes to the current feature subset. In a BF search, unlike greedy hill climbing, backtracking is allowed. That is, while exploring the feature space, if local changes to the current feature subset begin to look less promising, a BF search can backtrack to a more promising previous subset of features and continues the search from there. In other words, without any termination criterion, a BF search explores the entire feature space. Avoiding an endless loop in the reported work, the number of fully expanded subsets that lead to no improvement is limited to 5. The BF search algorithm is presented in Figure 3.

1. Begin with the OPEN list containing the start state, the CLOSED list empty, BEST ← start state.
2. Let s = arg max e(x) (get the state from OPEN with the highest evaluation).
3. Remove s from OPEN and add it to CLOSED.
4. If e(s) ≥ e(BEST), then BEST ← s.
5. For each child t of s that is not in the OPEN or CLOSED list, evaluate and add to OPEN.
6. If BEST changed in the last set of expansions, go to 2.
7. Return BEST.

Figure 3. The best-first (BF) search algorithm.

2.5. Experimental setup

In the reported work, the investigated engine defect is in the ignition system, i.e. the engines are operating with only the first cylinder missing firing. The engine sound signals are recorded in the workshop using a microphone located 20 cm above the engine. The investigated engines are from 4 different automobiles, including the Pride (Kia motors), the Peugeot 405, the Peugeot Pars, and Iran Khodro's national automobile, the Samand. For each automobile, the sound signals of 60 sampled automobiles in both healthy and faulty conditions with engines operating at 1000 rpm are recorded. There is only one faulty and one healthy state in the experimental dataset (Figure 4).

The sampling frequency used for the data acquisition was 44100 Hz. The whole process is depicted in Figure 5.

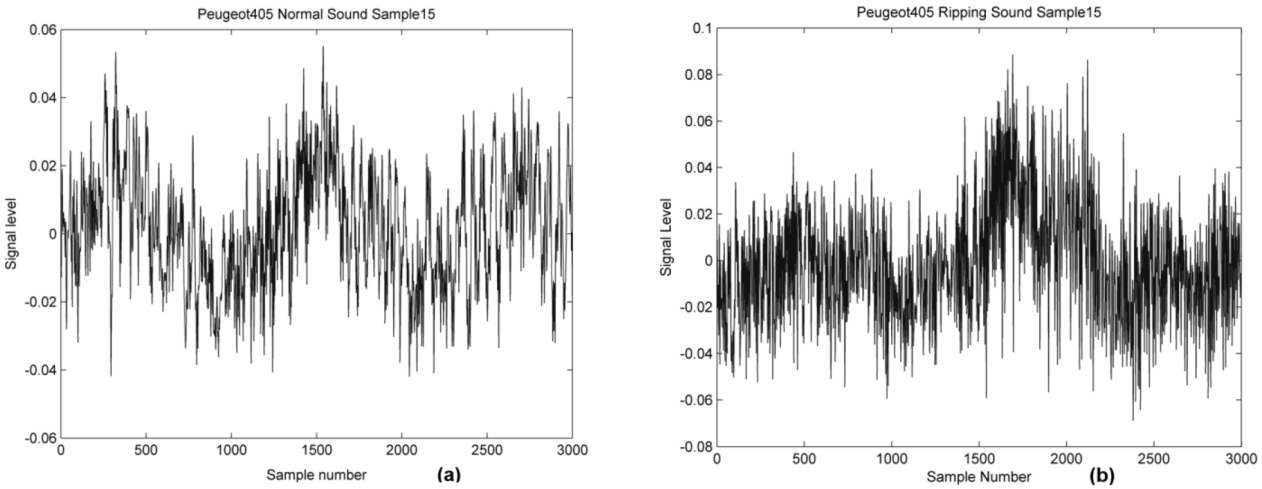

Figure 4. a) sample signal for the sound of normal operation of engine, b) sample signal for the sound of engine when it is ripping.

The recorded sound signals are transferred to the preprocessing unit where the recorded signals are manually denoised. Denoising is performed by listening to the recorded sound signals and separating parts of signals free from manmade inferred noise. Because the signals are recorded in the workshop, the sounds of other

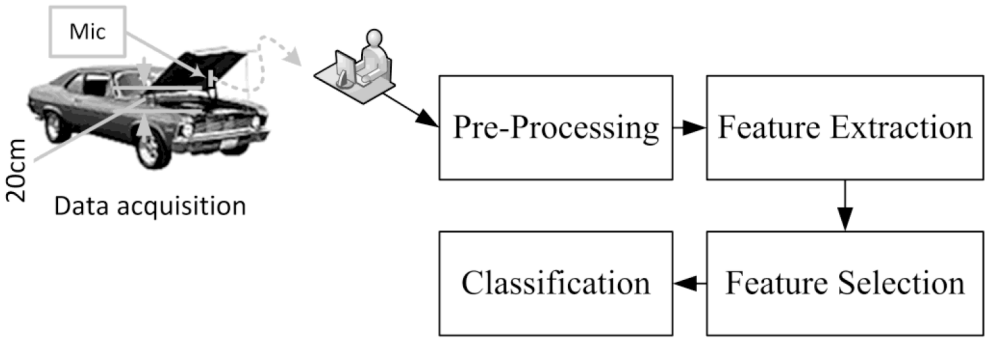

Figure 5. The whole process for acoustic sample collection, feature extraction, feature selection, and classification.

equipment in operation near the test subject and human voices are considered environmental noise. The signal level of the noise is so high that it reduces the signal to noise ratio to an unacceptably low level. As automobiles are being examined by repairmen in the workshop, the automobiles may or may not experience other additional faults. For example, the combustion-timing defect causes the engine to not operate properly, and this results in considerable sound abnormality. The analyzed signals include both healthy and faulty operating conditions at a recording time of 5 s.

In earlier reported works, the authors used the above-mentioned data using FFT [42] and DWT [43] approaches to distinguish faulty engines from healthy ones. Using the aforementioned data in the current work, the extracted signals are transferred to the feature extraction unit where each sound signal is decomposed by a WPT into a 3-level wavelet packet tree. Each node at the last level of the decomposition tree is used to calculate the aforementioned features in Section 2.1, and a dataset with 32 features is then constructed. Experiments show that moving from level 2 to level 3 did not make significant change in the selected features. Hence, we did not go any further. In the next step, the constructed dataset is transferred to the feature selection unit. In the feature selection unit, a CFS algorithm is used to select the most informative features while other, less informative ones are discarded. The reduced dataset is then transferred to the classification unit where 10% of the sample population is randomly selected for the training and 90% for the test. The classification techniques used in the reported work are the K-nearest neighbor (KNN) technique with a parameter of K = 5, the support vector machine (SVM) technique with a radial-basis function (RBF) kernel function, and the multilayer perceptron (MLP) technique. In the reported paper, the parameter (σ) for the RBF kernel function is 1 and the MLP used a back propagation (BP) training algorithm, a learning rate equal to 0.3, and momentum equal to 0.2. Experiments show that the engine problem is mostly in a [8268.75–19293.75] Hz frequency range, and we noted that other frequencies are less important and could be filtered. However, in different operation conditions, the frequency boundary could be adjusted for different categories of signal sources to gain better results.

3. Results and discussion

As the first step, the datasets made by a large number of wavelet bases using the aforementioned features are used for classification. Later on, using the CFS algorithm, the dimensionality of the sampled datasets is reduced (the most informative features are selected), and then each reduced dataset is used for classification. Table 3 introduces the final feature set and their feature numbers. Table 4 presents features selected for each constructed dataset from the wavelet bases. The reason why Energy, Kurtosis, Min, and Max features are selected is due to the better experimental results in signal decomposition using wavelet transformation. In other words, the

frequency of the appearance in the list of the selected features after feature selection is higher than the rest of the features. The energy feature provided good results with the highest frequency resolution. Results show that as the selected features are mostly from higher frequencies (signal details), it can be said that higher frequencies can be used to better separate different operation conditions of the engine. Figures 6 to 11 show the true positive rate (TPR) vs. the false positive rate (FPR) and the accuracy of each classification technique before and after dimensionality reduction on the dataset. It should be mentioned that the class of signals emitted from faulty engines are considered positive. Furthermore, the TPR, FPR, and accuracy are calculated using Eqs. (39–41), respectively.

Table 3. Features and their numbers in the dataset.

Feature numbers	Features
1 (AAA), 5 (AAD), 9 (ADD), 13 (ADA), 17 (DDD), 21 (DDA), 25 (DAA), 29 (DAD)	Max
2 (AAA), 6 (AAD), 10 (ADD), 14 (ADA), 18 (DDD), 22 (DDA), 26 (DAA), 30 (DAD)	Min
3 (AAA), 7 (AAD), 11 (ADD), 15 (ADA), 19 (DDD), 23 (DDA), 27 (DAA), 31 (DAD)	Kurtosis
4 (AAA), 8 (AAD), 12 (ADD), 16 (ADA), 20 (DDD), 24 (DDA), 28 (DAA), 32 (DAD)	Mean

Table 4. Selected features of each dataset after applying a CFS algorithm.

Wavelet base	No. of features	Selected features	Wavelet base	No. of	Selected features
coif 1	4	27,28,30,31	db 19	4	12,14,17,21
coif 2	12	3,10-12,14,17,19, 22,27,28,30,31	db 20	4	12,14,17,21
coif 3	2	14,17	sym 1	5	26-28,30,32
coif 4	3	14,17,19	sym 2	5	26-28,30,31
coif 5	3	14,17,25	sym 3	8	4,12,22,26-28,30,31
db 1	5	26–28,30,32	sym 4	11	3,10-12,14,15, 17,22,25,28,30
db 2	5	26–28,30,31	sym 5	3	12,14,17
db 3	8	4,12,22,26–28,30,31	sym 6	2	14,17
db 4	8	11,17,22,26-28,30,32	sym 7	3	14,17,25
db 5	4	11,14,17,27	sym 8	3	14,17,25
db 6	3	11,14,17	sym 9	3	14,17,32
db 7	3	12,17,22	sym 10	3	9,14,17
db 8	3	14,17,25	sym 11	3	14,17,32
db 9	3	14,17,25	sym 12	4	14,17,21,28
db 10	3	8,14,17	sym 13	4	14,17,21,32
db 11	3	14,17,25	sym 14	4	12,14,17,21
db 12	4	10,14,17,21	sym 15	5	10,14,17,21,31
db 13	4	12,14,17,21	sym 16	4	14,17,21,27
db 14	4	11,14,17,21	sym 17	4	11,14,17,21
db 15	4	12,14,17,21	sym 18	4	12,14,17,21
db 16	4	14,17,21,28	sym 19	5	14,17,20,21,27
db 17	4	12,14,17,21	sym 20	4	12,14,17,21
db 18	4	14,17,21,28			

$$True\,Positive\,Rate\,(TPR) \;=\; \frac{TP}{TP \;+\; FN} \tag{40}$$

$$False\,Positive\,Rate\,(FPR) \;=\; \frac{FP}{FP \;+\; TN} \tag{41}$$

$$Accuracy \;=\; \frac{TP \;+\; TN}{TP \;+\; TN \;+\; FP \;+\; FN}, \tag{42}$$

where TP, TN, FP, and FN are described as follows:

TP: The number of true instances that are correctly classified.

TN: The number of negative instances that are correctly classified.

FP: The number of negative instances that are incorrectly classified.

FN: The number of positive instances that are incorrectly classified.

First, it should be mentioned that the abbreviations BR and AR stand for before and after reduction, respectively. Looking at Figures 6 to 11, it is obvious that for all the constructed datasets, the classification results improve once the CFS feature selection algorithm is used. According to Figures 6 to 11, the Coiflet wavelet base with orders of 1 and 2, the Daubechies wavelet base with order 1–5, and Symlet wavelet base with an order of 1–4 does not lead to good classification results. For the other datasets in Table 4, it is clear that the smallest number of dataset dimensions resulted from sym6 and coif3, i.e. the corresponding datasets contain only 2 features: features 14 and 17. The highest ACC for dataset constructed from coif3 is 96.53%, a result of the KNN classification method, and the highest ACC for the dataset constructed from sym6 is 95.37%, which is a result of the KNN classification method. There are other datasets with 3 features, i.e. datasets constructed from coif4, coif5, db6–db11, sym5, and sym7–sym11. Among those datasets, the dataset constructed from coif4 has the highest ACC at 96.99% with KNN. Datasets with 4 features are also constructed by other wavelet bases, i.e. db5, db12–db20, sym12–sym14, sym16-sym18, and sym20. Among them, the dataset constructed from sym17 has the highest ACC at 96.99% with KNN. Considering the results in terms of classification accuracy, the best result is obtained using coif4. The corresponding features were 14, 17, and 19, corresponding to the kurtosis of packet 4 (ADA), the energy of packet 5 (DDD), and the maximum of packet 5 (DDD), respectively.

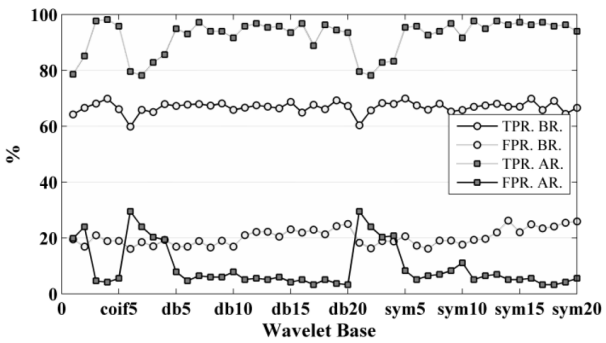

Figure 6. Comparing TRP vs. FPR before and after CFS for KNN.

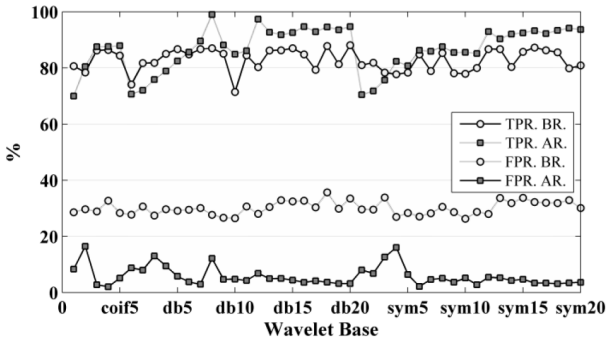

Figure 7. Comparing TRP vs. FPR before and after CFS for SVM.

Table 5 shows the confusion matrices for classification of datasets of coif3, coif4, sym6, and sym17 for all classification methods. Searching Table 4 for the features selected by CFS from the datasets of coif3, coif4, sym6, and sym17, one can see that the kurtosis of packet 4 and the energy of packet 5 are the most important features. Packets 4 and 5 correspond to the frequencies in the range of [8268.75–11025] Hz and [16537.5–19293.75] Hz,

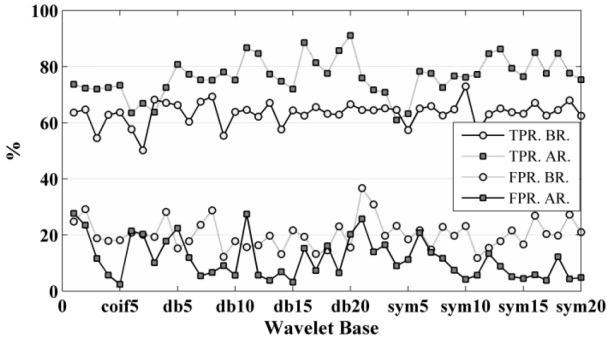

Figure 8. Comparing TRP vs. FPR before and after CFS for MLP.

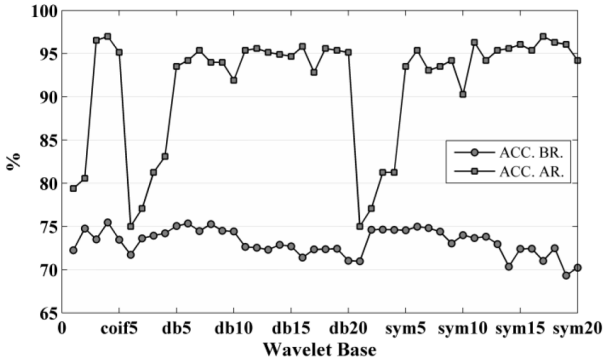

Figure 9. Accuracy before and after CFS for KNN.

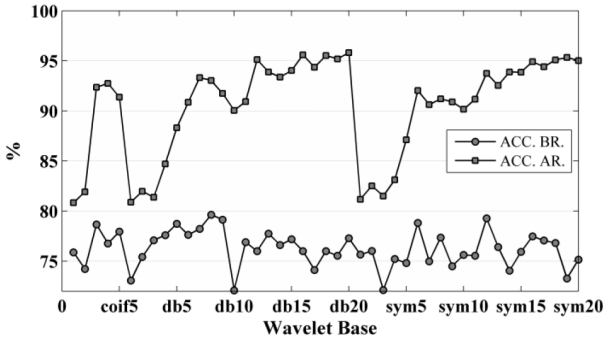

Figure 10. Accuracy before and after CFS for SVM.

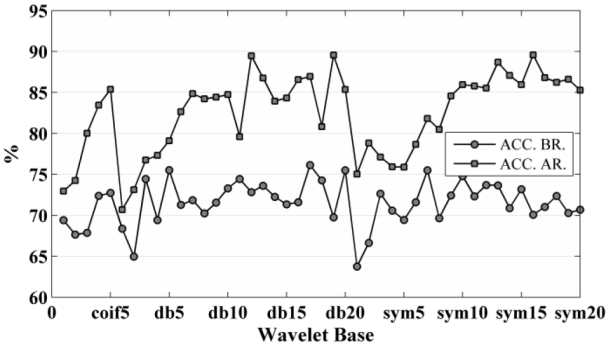

Figure 11. Accuracy before and after CFS for MLP.

respectively. Therefore, it can be claimed that the investigated fault affects the information of sound signals at a frequency range of [8268.75–19293.75] Hz more noticeably than the other frequencies (Figure 12).

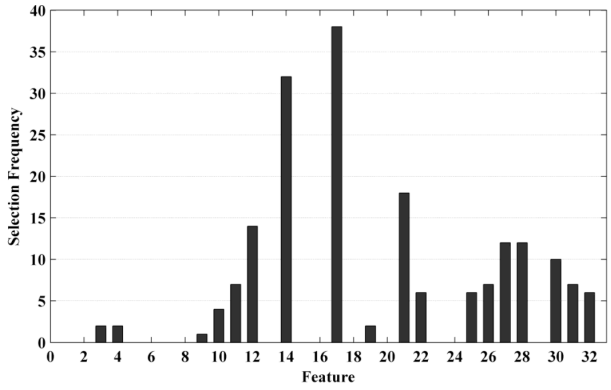

Figure 12. Feature frequency distribution.

In the previous works mentioned, the authors used both FFTs [42] and DWTs [43] to distinguish faulty engines from healthy ones. Using an FFT, the authors segmented the frequency spectrum to 9 different segments. Using the energy of the absolute value of FFT coefficients for each segment as signal signatures and PCA for feature reduction, they reached their best result with a 1500 Hz frequency segmentation with 83.45% accuracy.

For the DWT approach, the authors decomposed signals to 3 decomposition levels. Using the energy, kurtosis, variance, mean, minimum, and maximum of the wavelet coefficients for both approximation and

Table 5. Confusion matrix for classification of dataset of coif3, coif4, sym6, and sym17.

coif3								
KNN			SVM			MLP		
Classified As			Classified As			Classified As		
True Class	Faulty	Healthy	True Class	Faulty	Healthy	True Class	Faulty	Healthy
Faulty	210	5	Faulty	197	18	Faulty	178	37
Healthy	10	207	Healthy	17	200	Healthy	44	173
Classification Results			Classification Results			Classification Results		
ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)
96.53	97.67	4.61	91.9	91.63	7.83	81.37	75.83	13.02
coif4								
KNN			SVM			MLP		
Classified As			Classified As			Classified As		
True Class	Faulty	Healthy	True Class	Faulty	Healthy	True Class	Faulty	Healthy
Faulty	211	4	Faulty	191	24	Faulty	140	76
Healthy	9	208	Healthy	7	210	Healthy	11	205
Classification Results			Classification Results			Classification Results		
ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)
96.99	98.14	4.15	92.82	88.84	3.23	83.77	72.57	4.92
sym6								
KNN			SVM			MLP		
Classified As			Classified As			Classified As		
True Class	Faulty	Healthy	True Class	Faulty	Healthy	True Class	Faulty	Healthy
Faulty	206	9	Faulty	197	18	Faulty	169	47
Healthy	11	206	Healthy	17	200	Healthy	45	171
Classification Results			Classification Results			Classification Results		
ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)
95.73	95.81	5.07	92.04	86.3	2.17	78.66	78.28	20.78
sym17								
KNN			SVM			MLP		
Classified As			Classified As			Classified As		
True Class	Faulty	Healthy	True Class	Faulty	Healthy	True Class	Faulty	Healthy
Faulty	209	6	Faulty	197	18	Faulty	168	48
Healthy	7	210	Healthy	17	200	Healthy	8	208
Classification Results			Classification Results			Classification Results		
ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)	ACC (%)	TPR (%)	FPR (%)
96.99	97.21	3.23	94.4	92.19	3.37	86.81	77.77	3.7

detailed versions for each decomposition level as the signal signature and CFS features selection algorithm, they reached their best result for the first decomposition level with 80.67% accuracy. Analyzing the results, the authors reached the conclusion that the reason for the FFT outperforming the DWT approach is the fact that in the FFT approach features are extracted from high-frequency harmonics. However, in the DWT approach, features are extracted from low-frequency components. Consequently, the decision was made to choose between low- and high-frequency components of the signal. Thus, the WPT was selected as the alternative solution. Applying the selected approach, the results show that the WPT outperforms the aforementioned methods.

4. Conclusion

This paper reported on a study in which sound signals are analyzed using a WPT to identify faulty combustion of an automobile engine, regardless of the type of automobile. The proposed method is capable of dealing with the signals that contain environmental noises and are static during the recording time, e.g., the sound of a fan in operation. The proposed method is also applicable for other types of automobile engines. This was proven by the use of different types of automobiles as test subjects. Therefore, the proposed method proves to be automobile independent in its fault detection. Thus, among different signal analysis techniques in both frequency and time-frequency domains, the WPT has proven to be more successful. In the frequency domain signal processing, there is no time information. Hence, in some way, time information should be included in the calculations. The traditional WT does not represent high frequency information in detail. Therefore, because the WPT treats both high and low frequency in the same way, the information extracted from WPT coefficients can represent a more accurate signature. In conclusion, the best classification method is considered to be KNN.

5. Future work

In terms of future work, detecting a specific automobile or categorized detection of similar automobiles could be considered. Adding more faults to the list of the faults and their successful classification will also be included in future plans for the reported work. Automated denoising of the sound signals is a necessary process that will help with the automation of the system as a whole and can also be considered something to look at in the future. Since the preparation of the current text, the research has already improved. The use of PCA dimensionality of the input features to the classifier has been reduced, and the performance of the method has improved. This recent information is included in the MSc thesis of one of the authors and may be used in future publications.

Acknowledgments

The authors would like to thank the Irankhodro Powertrain Company (IPCO), a subsidiary of the Iran Khodro Company, which is a leading Iranian automaker (Mr. Izanloo), and Iran Khodro Central Repair Shop Number 5 (Mr. Saghi), which supported this work by giving us access to their facilities to collect samples.

References

[1] Lin J, Qu L. Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis. J Sound Vib 2000; 234: 135-148.

[2] Xian GM, Zeng BQ. An intelligent fault diagnosis method based on wavelet packer analysis and hybrid support vector machines. J Expert Syst Appl 2009; 36: 12131-12136.

[3] Albarbar A, Gu F, Ball AD. Diesel engine fuel injection monitoring using acoustic measurements and independent component analysis. J Measurement 2010; 43: 1376-1386.

[4] Wu JD, Liu CH. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. J Expert Syst Appl 2009; 36: 4278-4286.

[5] Figlus T, Liščák Š, Wilk A, Łazarz B. Condition monitoring of engine timing system by using wavelet packet decomposition of a acoustic signal. J Mech Sci Technol 2014; 28: 1663-1671.

[6] Wang YS, Ma QH, Zhu Q, Liu XT, Zhao LH. An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine. J Appl Acoust 2014; 75: 1-9.

[7] Jena DP, Panigrahi SN. Motor bike piston-bore fault identification from engine noise signature analysis. J Appl Acoust 2014; 76: 35-47.

[8] Wu JD, Chiang PH, Chang YW, Shiao YJ. An expert system for fault diagnosis in internal combustion engines using probability neural network. J Expert Syst Appl 2008; 34: 2704-2713.

[9] Al Kazzaz SAS, Singh GK. Experimental investigations on induction machine condition monitoring and fault diagnosis using digital signal processing techniques. J Electr Pow Syst Res 2003; 65: 197-221.

[10] El Hachemi Benbouzid M. A review of induction motors signature analysis as a medium for faults detection. IEEE T Ind Electron 2000; 47: 984-993.

[11] Junsheng C, Dejie Y, Yu Y. Time–energy density analysis based on wavelet transform. NDT&E Int 2005; 38: 569-572.

[12] Wu JD, Chen JC. Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines. NDT&E Int 2006; 39: 304-311.

[13] Zheng H, Li Z, Chen X. Gear fault diagnosis based on continuous wavelet transform. Mech Syst Signal Pr 2002; 16: 447-457.

[14] Saravanan N, Ramachandran KI. Fault diagnosis of spur bevel gear box using discrete wavelet features and decision tree classification. J Expert Syst Appl 2009; 36: 9564-9573.

[15] Wu JD, Liu CH. Investigation of engine fault diagnosis using discrete wavelet transform and neural network. J Expert Syst Appl 2008; 35: 1200-1213.

[16] Widodo A, Yang BS, Han T. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. J Expert Syst Appl 2007; 32: 299-312.

[17] Zuo MJ, Lin J, Fan X. Feature separation using ICA for a one-dimensional time series and its application in fault detection. J Sound Vib 2005; 287: 614-624.

[18] Saravanan N, Kumar Siddabattuni VNS, Ramachandran KI. A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box. J Expert Syst Appl 2008; 35: 1351-1366.

[19] Li W, Parkin RM, Coy J, Gu F. Acoustic based condition monitoring of a diesel engine using self-organising map networks. J Appl Acoust 2002; 63: 699-711.

[20] Du Z, Jin X, Wu L. Fault detection and diagnosis based on improved PCA with JAA method in VAV systems. J Build Environ 2007; 42: 3221-3232.

[21] Jack LB, Nandi AK. Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. Mech Syst Signal Pr 2002; 16: 373-390.

[22] Samanta B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. J Mech Syst Signal Pr 2004; 18: 625-644.

[23] Samanta B, Al-Balushi KR, Al-Araimi SA. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. J Eng Appl Artif Intel 2003; 16: 657-665.

[24] Elamin F, Gu F, Ball A. Diesel engine injector faults detection using acoustic emissions technique. Modern Applied Science 2010; 4: 3-13.

[25] Tse PW, Yang WX, Tam HY. Machine fault diagnosis through an effective exact wavelet analysis. J Sound Vib 2004; 277: 1005-1024.

[26] Vong CM, Wong PK. Engine ignition signal diagnosis with wavelet packet transform and multi-class least squares support vector machines. J Expert Syst Appl 2011; 38: 8563-8570.

[27] Vong CM, Wong PK, Ip WF. Case-based expert system using wavelet packet transform and kernel-based feature manipulation for engine ignition system diagnosis. J Eng Appl Artif Intel 2011; 24: 1281-1294.

[28] Nikolaou NG, Antoniadis IA. Rolling element bearing fault diagnosis using wavelet packets. NDT&E Int 2002; 35: 197-205.

[29] Zarei J, Poshtan J. Bearing fault detection using wavelet packet transform of induction motor stator current. Tribol Int 2007; 40: 763-769.

[30] Fan X, Ming JZ. Gearbox fault detection using Hilbert and wavelet packet transform. Mech Syst Signal Pr 2006; 20: 966-982.

[31] Jensen A, Cour-Harbo A. la. Ripples in Mathematics: The Discrete Wavelet Transform. 1st ed. Berlin, Germany: Springer, 2001.

[32] Hu X, Wang Z, Ren X. Classification of surface EMG signal using relative wavelet packet energy. J Comput Meth Prog Bio 2005; 79: 189-195.

[33] Daubechies I. Orthagonal bases of compactly supported wavelets. Comm Pure Appl Math 1988; 41: 909-996.

[34] Härdle W, Kerkyacharian G, Picard, D, TsybakovA. Wavelets, Approximation, and Statistical Applications. 1st ed. New York, NY, USA: Springer 1998.

[35] Siddiqi AH. Applied Functional Analysis: Numerical Methods, Wavelet Methods, and Image Processing. Boca Raton, FL, USA; CRC Press, 2003.

[36] Daubechies I. Ten Lectures on Wavelets. Philadelphia, PA, USA: SIAM, 1992.

[37] Beylkin G, Coifman R, Rokhlin V. Fast wavelet transforms and numerical algorithms I. Comm. Pure Appl. Math 1991; 44: 141-183.

[38] Monzón L, Beylkin G. Compactly supported wavelets based on almost interpolating and nearly linear phase filters (coiflets). Appl Comput Harmon A 1999; 7: 184-210.

[39] Fayyad UM, Irani KB. Multi-interval discretization of continuous-valued attributes for classification learning. In: 13th International Joint Conference on Artificial Intelligence (IJCAI); 28 August–3 September 1993; Beijing, China: Advancement of Artificial Intelligence Press. pp. 1022-1027.

[40] Hall MA, Smith LA. Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In: Twelfth International FLAIRS Conference; 1–5 May 1999; Orlando, FL, USA: Advancement of Artificial Intelligence Press. pp. 235-239.

[41] Quinlan JR. Induction of decision trees. J Mach Learn 1986; 1: 81-106.

[42] Ghaderi H, Kabiri P. Automobile independent fault detection based on acoustic emission using FFT. In: Singapore International Non-Destructive Testing Conference & Exhibition; 3–4 November 2011; Singapore.

[43] Ghaderi H, Kabiri P. Automobile independent fault detection based on acoustic emission using wavelet. In: Singapore International Non-Destructive Testing Conference & Exhibition; 3–4 November 2011; Singapore.