Cognitive Computation

A Novel Multiple Feature-based Engine Knock Detection System using Sparse Bayesian Extreme Learning Machine

--Manuscript Draft--

Manuscript Number:	COGN-D-20-00117R3
Full Title:	A Novel Multiple Feature-based Engine Knock Detection System using Sparse Bayesian Extreme Learning Machine
Article Type:	Original Article
Keywords:	Engine Knock Detection; Variational Mode Decomposition; Multiple Feature Learning; sample entropy; Sparse Bayesian Extreme Learning Machine
Corresponding Author:	Pak Kin Wong, Ph.D. University of Macau Taipa, MACAO
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Macau
Corresponding Author's Secondary Institution:	
First Author:	Zhaoxu Yang, PhD
First Author Secondary Information:	
Order of Authors:	Zhaoxu Yang, PhD
	Haijun Rong, PhD
	Pak Kin Wong, Ph.D.
	Plamen Angelov, PhD
	Chi Man Vong, PhD
	Chi Wai Chiu, MSc.
	Zhixin Yang, PhD
Order of Authors Secondary Information:	
Funding Information:	

Funder	Recipient
Fundo para o Desenvolvimento das Ciências e da Tecnologia (0021/2019/A, 0018/2019/AKP and 0008/2019/AGJ)	Not applicable
National Natural Science Foundation of China (61976172)	Prof. Haijun Rong
Natural Science Basic Research Program of Shaanxi Province (020JQ-013)	Dr. Zhaoxu Yang
Macao Youth Scholars Program (AM201909)	Dr. Zhaoxu Yang
University of Macau Distinguished Visiting Scholar Program	Not applicable
Natural Science Basic Research Program of Shaanxi Province (2020JM-072)	Prof. Haijun Rong
National Natural Science Foundation of China (12002254)	Dr. Zhaoxu Yang
Multi-Year Research Grant from the University of Macau (MYRG2019-00137-FST)	Not applicable

Abstract:	Background Automotive engine knock is an abnormal combustion phenomenon that affects engine performance and lifetime expectancy, but it is difficult to detect. Collecting engine vibration signals from an engine cylinder block is an effective way to detect engine knock. Methods This paper proposes an intelligent engine knock detection system based on engine vibration signals. First, filtered signals are obtained utilizing variational mode decomposition (VMD), which decomposes the original time domain signals into a series of intrinsic mode functions (IMFs). Moreover, the values of the balancing parameter and the number of IMF modes are optimized using genetic algorithm (GA). IMFs with sample entropy higher than the mean are then selected as sensitive subcomponents for signal reconstruction and subsequently removed. A multiple feature learning approach that considers time domain statistical analysis (TDSA), multi-fractal detrended fluctuation analysis (MFDFA) and alpha stable distribution (ASD) simultaneously is utilized to extract features from the denoised signals. Finally, the extracted features are trained by sparse Bayesian extreme learning machine (SBELM) to overcome the sensitivity issue of hyperparameters in conventional machine learning algorithms. Results A test rig is designed to collect the raw engine data. Compared with other technology combinations, the optimal scheme from signal processing to feature classification is obtained, and the classification accuracy of the proposed integrated engine knock detection method can achieve 98.27%. Conclusions We successfully propose and test a universal intelligence solution for the detection task.
Additional Information:	
Question	Response
What is the main contribution of this paper to the Cognitive Computation community - in a couple of sentences, including the 'cognitive' or 'biologically- inspired' computational aspects of your work - this should also be appropriately highlighted in the paper (including in the abstract, introduction etc)?	A new intelligent engine knock detection system using multiple feature based-sparse Bayesian extreme learning machine, genetic algorithm-based signal processing method & sample entropy is proposed.
Why is this contribution significant (what impact will it have)?	The study can detect the engine knock accurately and hence reduce the chance of engine failure. This is also the first research on using advanced machine learning approach for engine knock detection.
Please identify a few (most recent) papers that are closely related to your work. Be sure that these comparisons are appropriately introduced in your work and be sure to critically review other related cognitively/biologically-inspired works recently reported in current research literature.	In recent, there is no any related paper in Congnitive Computation.
What is distinctive/new about the current paper relative to the above (previously published) works (and also relative to authors' own related/previous works, if any)? (please ensure this novelty is appropriately highlighted in your submitted paper, including in the title, abstract, and elsewhere, as appropriate)	The existing studies on engine knock detection use one kind of feature from engine. This work is the first attempt at using multiple features form engine to train a parameter-insensitive classifier.

COGNITIVE COMPUTATION

Manuscript No.COGN-D-20-00117R2 Summary of Changes and Replies to Comments of the Associate Editor and Referees

2nd September, 2021

We would like to thank the editor and reviewers for their insightful comments and invaluable help on our submission. We have taken the comments very seriously and modified the paper based on the suggestions provided.

I. Comments of Editor

Whilst the reviewers are happy with the technical revisions, the paper's presentation is still below par for publication in this journal (a number of sentences throughout are overly long/vague/unclear/confusing, with mixed/improper use of tenses, grammar etc.). Authors are required to carefully proof read the paper (with help from an experienced, native English speaker, or preferably a professional service), or paper may be rejected in the next final re-submission opportunity.

Reply: Thank you for your comment. We have recognized that there are many problems existed in the paper presentation. We have used the language editing service support by Springer Nature Author Services (SNAS), and obtained professional proof reading, including correction of English language, grammar, punctuation, spelling, and overall style. The SNAS provided detailed suggestions in the "Completed Files.docx" and the "Editing Summary.PDF". In addition, a Editing Certificate was attached. According to the revision suggestions, we have rewritten and proofread the submission.

Editing Certificate

This document certifies that the manuscript

A Novel Multiple Feature-based Engine Knock Detection System using Sparse Bayesian Extreme Learning Machine

prepared by the authors

Zhao-Xu Yang, Hai-Jun Rong, Pak Kin Wong, Plamen Angelov, Chi Man Vong, Chi Wai Chiu, and Zhi-Xin Yang

was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at SNAS.

This certificate was issued on **September 7, 2021** and may be verified on the SNAS website using the verification code **9A59-102C-67DC-9A80-70FP** .

Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes. To verify the final SNAS edited version, please visit our verification page at secure.authorservices.springernature.com/certificate/verify.
If you have any questions or concerns about this edited document, please contact SNAS at support@as.springernature.com.

September 07, 2021

Dear Zhaoxu Yang,

Thank you for choosing Springer Nature Author Services. This manuscript, titled "A Novel Multiple Feature-based Engine Knock Detection System using a Sparse Bayesian Extreme Learning Machine," is very interesting. The paper was edited for grammar, phrasing, and punctuation. In addition, many edits were made to further improve the flow and readability of the text. Below, we highlight the areas of this paper that we focused on in our edit.

Articles are an important aspect of the English language, including the definite article "the" and the indefinite articles "a" and "an." Our edits focused on improving article use, which is often strongly dependent on context and field conventions.

The easiest way to avoid using vague pronouns in your writing is to use demonstrative pronouns as adjectives that modify a more descriptive term (e.g., "This inconsistency" or "These findings") and to replace pronouns such as "It" with more specific nouns.

Certain edits were made to remove redundant, repetitive or unnecessary phrasing and to present the information in a more straightforward manner.

Comments were left if further clarification would be helpful or confirmation of the meaning of the text was necessary. Please review these comments and all our changes carefully for more detailed suggestions, as well as to ensure that the final version of the manuscript is fully accurate.

Thank you again for using our editing services; we wish you the best of luck with your submission.

Best regards,

Kim P.
Senior Editor
Springer Nature Author Services

A Novel Multiple Feature-based Engine Knock Detection System using a Sparse Bayesian Extreme Learning Machine

Zhao-Xu Yang, Hai-Jun Rong, Pak Kin Wong, Plamen Angelov, Chi Man Vong, Chi Wai Chiu, and Zhi-Xin Yang

Abstract.

Background ~~The automotive~~ Automotive engine knock is an abnormal combustion phenomenon ~~which~~that affects ~~the~~ engine performance and lifetime expectancy, but it is difficult to detect. Collecting e~~E~~ngine vibration signals ~~collected~~ from ~~the~~ an engine cylinder block is an effective way to detect engine knock.

Methods This paper proposes an intelligent~~ce~~ engine knock detection system based on engine vibration ~~signal. Firstly~~signals. First, filtered signals are obtained by utilizing ~~the~~ variational mode decomposition (VMD), which decomposes the original time domain signals into a series of intrinsic mode functions (IMFs). Moreover, the values of the balancing parameter and the number of IMF modes ~~of IMFs~~ are optimized using a genetic algorithm (GA). IMFs with sample entropy higher than the mean are then selected as sensitive subcomponents for signal reconstruction and subsequently removed. A multiple feature learning approach ~~which~~that considers time domain statistical analysis (TDSA), ~~multi-fractal~~multifractal detrended fluctuation analysis (MFDFA) and the alpha stable distribution (ASD) simultaneously~~,~~ is utilized to extract features from the denoised signals. Finally, the extracted features are trained by a sparse Bayesian extreme learning machine (SBELM) to overcome the sensitivity~~e~~ issue of hyperparameters in conventional machine learning algorithms. **Results** A test rig is designed to collect the raw engine data. Compared with other~~a~~ ~~the mass of combination~~ technology combinations~~ies involved~~, the optimal scheme from signal processing to feature classification is obtained, and the classification accuracy of the proposed integrated engine knock detection method can achieve 98.27%~~ in engine knock detection~~. **Conclusions** We successfully propose and test a universal intelligence solution for the detection task.

Keywords: Engine Knock Detection, Variational Mode Decomposition, Multiple Feature Learning, Sample Entropy, Sparse Bayesian Extreme Learning Machine.

1 Introduction

In a spark-ignition automotive engine, engine knock is defined as an abnormal combustion phenomenon ~~which~~that is observed as a ~~source of~~ noise and~~or~~ can even indicate a major engine fault. Engine k~~K~~nock is an essential factor ~~constraining~~ that

批注 [Ed2]: Please ensure that the intended meaning has been maintained in this edit.

带格式的: 字体: 加粗

constrains ~~the~~ further improvements ~~of~~in ~~the~~ thermal efficiency and causes serious engine damage, such as piston or cylinder breakage. During heavy knock, ~~a lot of~~ much extra heat is transferred to the combustion chamber wall, resulting in a rapid rise in the temperature ~~rise of~~ the piston and cylinder head ~~rapidly~~. The overheating of these parts makes the intensity of the knock continue to increase. This consequential runaway phenomenon may trigger ~~the~~ engine failure within a few minutes. Moreover, an excessively high pressure pulse due to heavy knock may occur in the end gas area. The interaction between this high local pressure and high local surface temperature inevitably ~~weaken~~weakens or ~~corrode~~corrodes the engine material.

On the premise of the accurate identification of signals associated with ~~the~~ engine knock, some preventive measures should be performed, such as delaying the ignition timing. These signals can be monitored and collected by pressure wave amplitude analysis, exhaust gas temperature analysis, heat transfer analysis, *etc.* [40]. However, the high cost of ~~the~~ in-cylinder pressure sensors, as well as the decreased lifetime expectancy resulting from ~~blends of hot contact~~high temperatures and high pressure, make the pressure wave amplitude method difficult to apply ~~be~~ extensively ~~popularized~~ [11]. Exhaust gas temperature analysis suffers from low precision [19], and heat transfer analysis is difficult to ~~obtain~~ apply in real time [20].

Massive pressure waves [21] occur inside of ~~the~~ an ignition chamber~~, which~~ and can emit an audible ~~hearable~~ sound, and the resulting vibrations create the perceptible knock signal. Therefore, engine vibration ~~signal is~~signals are widely used for engine knock detection, which is a compromised solution for resolving ~~conflict~~conflicts between ~~the~~ measuring precision and cost. When vibrations are~~is~~ detected in the cylinder wall, the knock sensor, which is a ~~crystal of the~~ piezoelectric crystal placed on ~~the~~ an engine cylinder block, creates a low voltage signal that is fed back to the electronic control unit (ECU). Knock can be determined when the resonant frequency is close to or beyond the frequency range of the knock frequency. However, ~~the~~ engine vibrations include~~s~~ not only ~~includes~~ the in-cylinder pressure pulse~~,~~ but also ~~the~~ piston slaps, valve train motion, fuel injector pulses and~~or~~ other engine structural ~~vibration~~vibrations, which have little influence on knock characteristics, but they ~~are easy to~~ easily conceal~~cover up~~ slight knock. Even though an advanced knock module can be installed to reduce ~~the~~ background ~~noises~~noise, the knock module requires expertise to tune the frequency band, central frequency and gains. In addition ~~fact~~, it is ~~also~~ difficult for ~~the expert~~ experts to determine the optimal parameters of the knock module to filter out ~~the~~ background noise under difficult time-varying conditions.

The vibration signal detection method uses an accelerometer to detect ~~the~~ knock characteristics by measuring the vibration acceleration of the cylinder block. Since this method has the advantages of easy installation, high reliability and low cost, it is commonly employed in real-time engine knock detection. Although using vibration signals to determine engine knock is more practical, a cylinder block vibration signal has a substantial amount of noise and signals from other vibration sources. Engine vibration signals ~~could not~~ cannot be applied to detect knock directly, and the original signals need to be processed using an accurate and effective signal denoising technique. ~~So~~Therefore, utilizing ~~the~~ vibration signals for engine knock detection is still a challenging task.

Engine knock detection is a complicated problem that includes~~ing~~ signal denoising, feature extraction, and feature classification. In ~~the~~ signal denoising, although signals using variational mode decomposition (VMD) [35] are separated into a series of intrinsic mode functions (IMFs), IMFs ~~depends~~depend on the values of the balancing parameter and the number of modes ~~which~~that are adjustable, and the results may be inaccurate when the parameters~~y~~ are not set in real time~~in place~~. ~~So~~Therefore, there is an urgent need ~~for obtaining~~to obtain the optimal values of the VMD parameters. To reduce the computational burden in the later stage, some nonlinear dynamic parameters, such as the energy ratio and correlation coefficient, should be ~~taken~~ used to extract the IMFs that represent prominent features. However, they are dependent on the record length, which ~~are~~is usually difficult or even impossible to ~~be acquired~~acquire, especially in online condition monitoring and diagnosis. ~~Appropriate~~An appropriate indicator is also needed to determine sensitive subcomponents to select and reconstruct important IMFs. During~~In the~~ feature extraction, each feature extraction method extracts different independent and complementary information from the signals~~ that are both independent and complementary~~. ~~So~~Therefore, an ensemble system using ~~the~~ multiple feature learning is proposed ~~in order~~ to achieve high classification accuracy. ~~An optimal feature combination usually needs a lot of~~ M~~m~~any experiments are usually needed to test the availability and performance of an optimal feature combination in specific ~~application~~applications. In feature classification, machine learning methods play an important role in the performance of the final classification results. Traditional neural networks and support vector machines have been applied to fault classification [17, 30]. However, they suffer from~~ the~~ issues, including the~~ of~~ computational burden of the large-scale fault classifier and the sensitivity of the~~e issue of~~ hyperparameters.

The main motivation of this research is to find the ~~most optimal~~ best solution in theory and application. In this paper, a novel intelligence engine knock detection system ~~by using~~using a multiple feature~~ based~~-based sparse Bayesian extreme learning machine (SBELM), genetic algorithm-based ~~variational mode decomposition~~VMD (GA-VMD) and sample entropy is proposed, and the salient contributions of this paper are organized as follows~~,~~:

1) The traditional engine knock detection system usually relies on one kind of feature extracted from engine vibrations. Considering that the combination of different feature spaces from the observations would ~~take on~~ achieve better performance than any base classifier, an ensemble system using~~ the~~ multiple feature learning is proposed ~~in order~~ to achieve high classification accuracy.
2) ~~In order to~~ To overcome the dependency of the appropriate values of the balancing parameter and the number of modes, GA-VMD is used to filter ~~the~~ unavoidable noise~~s~~, in which the genetic algorithm (GA) is applied to obtain the optimal parameters to enhance the noise reduction ability. When the original time domain

signals are decomposed into a series of IMFs, IMFs with sample entropy higher than the mean ~~is~~are selected as sensitive subcomponents for signal reconstruction.

3) This work is the first to attempt~~ at~~ applying multiple features captured ~~form~~from engine vibration ~~signal~~signals and SBELM together for engine knock detection. ~~Besides~~In addition to addressing the ~~computation~~computational burden issue of the large-scale fault classifier, the extracted features are trained by SBELM to overcome the sensitivity~~e issue~~ of hyperparameters in conventional machine learning algorithms.
4) A universal intelligence solution for the detection task~~,~~ and the integration of GA-VMD with~~,~~ sample entropy,~~ combined with~~ time domain statistical analysis (TDSA), ~~and~~ alpha stable distribution (ASD), and SBELM ~~is~~are also proposed to build an effective engine knock detection system.

This paper is organized as follows. The related work is briefly reviewed in Section 2. Section 3.1 introduces the outline of the engine knock detection system. The design procedure of the detection system ~~is presented together with~~ and the signal filtering method are presented in Section 3.2, the feature extraction technique is presented in Section 3.3, and the classification procedure, which involves multiple techniques, is presented in Section 3.4~~ that involve multiple techniques~~. The performance evaluations of the proposed detection system are given in Section 4. Finally, a conclusion is summarized in Section 5.

2 Related Work

We briefly review previous approaches related to engine knock detection.

2.1 Signal Denoising

Engine knock detection can be viewed as an engine fault detection problem~~,~~ that relies~~ying~~ on the features captured from a~~the~~ signal. The signal may contain ~~noises~~noise or it can be affected~~impact~~ by other component vibrations, so ~~that~~ the knock-related information contained therein is not easy to observe. Therefore, many efforts have been made to develop~~on~~ signal processing techniques [32], such as the fast Fourier transform [15], the short-time Fourier transform [27], the continuous wavelet transform [7,39], the discrete wavelet transform [6,34] and nonlinear wavelet transforms [16]. ~~Fast~~The fast Fourier transform method converts a time domain signal into a frequency domain signal quickly, but it is not suitable for ~~non-stationary~~nonstationary signals such as the knock signal, which experiences rapid changes in both time and frequency~~ rapidly~~. The s~~S~~hort-time Fourier transform is an alternative transform method~~ation~~ for time-frequency analysis, but it has not been extensively used due to its low time resolution with a fixed window under high frequencies. The resolution issue has been solved by wavelet transforms. However, the application of a wavelet transform has been bound by its inherent defect, which is the limitation of the selection of a mother wavelet, and it is~~indeed~~ a nona~~non-a~~daptive~~ly~~ transformation. Empirical mode decomposition (EMD) [31] is self-adaptable and decomposes a signal directly into several IMFs, which are defined as amplitude-modulated-frequency-modulated signals whose number of local extrema and zero-crossings differ at most by one [13]. For the phenomenon that mode mixing occurs repeatedly in EMD, ensemble ~~empirical mode decomposition~~EMD (EEMD), which ~~is proposed that~~ decreases the chance of undue mode mixing to a certain extent, was proposed [5]. The IMF in EEMD is characterized as the mean of an ensemble of trials whereby a finite-~~-~~amplitude white noise signal is added to the decomposed data in each trial;~~,~~ ~~with~~ this approach increases the~~ increase of~~ computational burden since the data size of IMF is equal to that of the raw data. In recent years, VMD has been introduced~~ into rotating machines~~ for noise analyse~~i~~s of rotating machines and as a fault diagnosis method ~~which~~that has shown very promising results [8,28,35]. Although signals are separated into a series of IMFs, IMFs ~~depends~~depend on the values of the balancing parameter and the number of modes ~~which~~that are adjustable, and the results may be inaccurate when they are not set in real time~~place~~ [3]. Therefore, an optimization method utilizing~~ the~~ GA is proposed in this work to solve the problem of parameter optimization.

To reduce the computational burden in the later stages, some nonlinear dynamic parameters, such as the energy ratio [38] and correlation coefficient, should be taken to extract the IMFs that represent prominent features. However, the reliable estimation of both parameters depends on very long datasets~~data sets~~, which are usually difficult or even impossible to ~~be acquired~~ acquire, especially during~~in~~ online condition monitoring and diagnosis. Entropy is defined as the loss of information in a time series or signal, such that approximate entropy [36] and sample entropy [24] are created to measure the repeatability or predictability within a time series. Due to its self-matching problem, approximate entropy is heavily dependent on the record length, and its value is uniformly lower than expected for short records~~,~~ and lacks relative coherence~~ as well~~. Sample entropy is less dependent on the time series length and ~~,~~ ~~which~~ is utilized in this work to select and reconstruct important IMFs.

2.2 Feature Extraction

The selection of the feature extraction algorithm is known to ~~be~~play an important role in determining the performance of the classification system. An ensemble system using ~~the~~ multiple feature learning is proposed ~~in order~~ to achieve high classification accuracy. This is made by combining the classifiers that are trained on different feature sets. The idea of combining different ~~features~~feature spaces from the observations made, that is, the combination of classifiers in different feature spaces, is the most effective way of combining classifiers and usually presents better results than any base classifier [9]. This occurs because each feature extraction method extracts different independent and complementary information from the signal~~ that are both independent and complementary~~. For this purpose, a diverse set of feature extraction

批注 **[Ed3]:** Please ensure that the intended meaning has been maintained in this edit.

methods using different approaches ~~like~~, such as TDSA, ASD and ~~multi-fractal~~multifractal detrended fluctuation analysis (MFDFA), are selected.

2.3 Feature Classification

After ~~the~~ feature exaction, machine learning methods play an important role in the performance of the final classification results. Traditional neural ~~network~~networks and support vector ~~machine were~~ machines have been applied to fault classification [10,14]. Much practical evidence shows that the long training time has greatly restricted the efficiency of these algorithms. In recent years, extreme learning ~~machine (ELM) is~~machines (ELMs) have been utilized for ~~multi-class~~multiclass classification based on a~~the~~ single hidden layer feed-forward network (SLFN) [12]. Recent studies show that the learning speed of ELM is faster than ~~the~~that of traditional learning algorithms [17,26], so ELM can be suitable~~competent~~ for large-scale problems~~-~~. The dependent parameter of ELM is the number of hidden ~~neurons~~neuron nodes, but the initial hidden node parameters are random. Considering the susceptibility caused by the number of hidden neurons in conventional ELM, there might be a large ~~amount~~number of hidden neurons selected in the trained model due to the minimization of the training error while~~in~~ ranking neurons, resulting in a ~~highly~~high computational cost. The SBELM classifier is presented in this work; it has ~~with~~ the benefits of a lower computational load ~~from~~ than ELM, ~~and~~ a small weight, and better ~~together with~~ prediction posterior probability than~~from~~ relevance vector machines (RVMs) [30]~~, which has~~. Hence, SBELM requires less calculation and is~~to be~~ more suitable as a large-scale fault classifier.

2.4 Previous Schemes

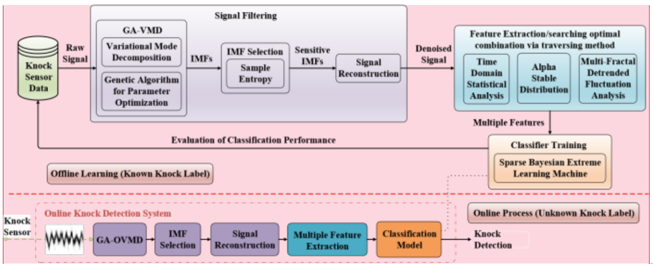


Fig. 1~~Fig.1~~. Engine knock detection framework and project work flow

Knock detection is usually a complicated problem ~~which~~that ~~needs to combine~~ requires a combination of multiple techniques. Some previous schemes provided~~gave~~ effective solutions by exploiting different technologies~~,~~ and ensured the reliability of

批注 **[Ed4]:** Please note that we have not checked the text in your images because they are not editable in Word. If you would like them to be edited, please contact our support staff.

~~the~~ knock detection. ~~A sound~~ Sound vibration signal processing was proposed in [25]. In [25], a combination of methods, such as pass high-frequency filters, normalized envelope functions and regression, were~~as~~ used to describe ~~the~~ knock patterns, and then, the Euclidean distance ~~gave~~ was used to determine ~~a decision on~~ the existence of a detonation and achieved an accuracy of ~~about~~approximately 95%. However, the ~~liner~~linear filter and distance-based classifier have limited ~~abilities of~~ noise reduction and feature classification abilities, respectively. A knock characteristic detection method based on wavelet~~-~~ denoising and EMD was proposed in [4]. The results indicated that the knock detection accuracy was 97%. ~~A~~An approach for detecting engine knocks ~~in~~of various intensities based on the vibration signal of an engine block using VMD and ~~semi-supervised~~semisupervised local ~~fisher~~Fisher discriminant analysis was proposed in [3], and the classification rate for~~of~~ strong knocks was over 95%. As mentioned above, there is much room for improvement in the denoising performance and accuracy.

3 Designed of the Engine Knock Detection System

3.1 Outline of the Detection System

Motivated by the above general engine fault diagnostic requirements, a novel practical engine knock detection framework and project ~~work flow~~ workflow are proposed in Fig. 1. The proposed framework contains three main sections~~, including~~ : signal filtering, feature extraction and classification. The GA-VMD method is developed to separate ~~noises~~noise from the raw signal with a low ~~computation~~computational burden compared with EEMD, where VMD is integrated with GA to achieve appropriate values of the balancing parameter and number of modes. While the VMD converts the original signal into a series of IMFs, sensitive IMFs are then selected by sample entropy for further filtered signal reconstruction, and th~~os~~e unconsidered IMFs are removed. In terms of candidate feature extraction techniques before fault classification, the TDSA, MFDFA and ASD methods~~,~~ and their possible combinations~~,~~ are tested to describe the distinguishable characteristics of the denoised signals~~, respectively~~. These features are trained by SBELM ~~for establishing~~to establish a precision classifier. After the features of an unseen signal are fed to the trained classifier, a universal detection scheme is achieved to ~~accurate~~accurately identify engine knock online, such that the ECU ~~could do~~can perform some actions ~~to protect the engine~~, such as the retardation of the ignition in advance, to protect the engine.

3.2 Signal Filtering

GA-VMD For ~~the~~ nonlinear and ~~non-stationary~~nonstationary time-~~—~~frequency ~~characteristic~~characteristics, GA-VMD is considered for signal filtering in the following work.

The goal of VMD is to decompose a real valued input signal *f* into a discrete number of subs~~sub-s~~ignals (i.e., IMFs) u_k~~,~~ that have specific sparsity properties while reproducing the input. Here, the sparsity property of each mode is chosen to be its bandwidth in the spectral domain. In other words, we assume the *k*th mode to be mostly compact around a center pulsation ω_k, which is to be determined along with the decomposition.

~~In order to~~ To assess the bandwidth of a mode, the following scheme is proposed. (i) For each mode u_k, the associated analytic signal is computed by means of the Hilbert transform ~~in order~~ to obtain a unilateral frequency spectrum. (ii) For each mode, the frequency spectrum of the mode is shifted to the baseband~~,~~ by mixing ~~with~~ an exponentially tuned value with~~to~~ the respective estimated center frequency. (iii) The bandwidth is now estimated through the Gaussian smoothness of the demodulated signal, *i.e.,* ~~i.e., .~~ the squared L_2–norm of the gradient. The resulting constrained variational problem is given as follows~~,~~:

$$\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \exp(-j\omega_k t) \right\|_2^2 \quad \text{s.t.} \sum_{k=1}^{K} u_k(t) = f(t) \tag{1}$$

where t is the time script, δ is the Dirac distribution and * denotes convolution. $\{u_k\} :=\{u_1,\ldots,u_K\}$ and $\{\omega_k\}:=\{\omega_1,\ldots,\omega_K\}$ are shorthand notations for the sets of all modes and their center frequencies, respectively. $k=1,2,\ldots, K$, and K is the number of modes of the intrinsic mode components.

The solution to Eq. (1) can be easily achieved via an unstrained optimization problem using the augmented Lagrangian method

$$\mathcal{L}(\{u_k\},\{\omega_k\},\lambda) := a\sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] \exp(-j\omega_k t) \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \tag{2}$$

where a is the balancing parameter of the data-fidelity constraint~~,~~ and λ is the Lagrange multiplier. An alternating direction method of multipliers is adopted to solve Eq. (2). The estimated modes u_k and the corresponding updated center frequency ω_k in the frequency domain can be achieved as follows:

$$u_k^{n+1}(\omega) = \frac{\tilde{f}(\omega) - \sum_{i<k} u_i^{n+1}(\omega) - \sum_{i>k} u_i^{n}(\omega) + \lambda^n(\omega)/2}{1 + 2a(\omega - \omega_k^n)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |u_k^{n+1}(\omega)|^2 d\omega} \tag{4}$$

where $\tilde{f}(\omega) := 1/\sqrt{2\pi}\int_{\mathbb{R}} f(t)\exp(-j\omega t)dt$ with $j^2 = -1$~~,~~ is the Fourier transform of the

批注 [Ed5]: Please ensure that the intended meaning has been maintained in this edit.

signal f(t). The Lagrangian multiplier is updated as:

$$\lambda^{n+1}(\omega) = \lambda^{n}(\omega) + \tau_0 \left(\tilde{f}(\omega) - \sum_k u_k^{n+1}(\omega) \right) \quad (5)$$

where τ_0 is the update parameter.

However, the values of the balancing parameter α and the number of modes K in Eq. (2) need to be predefined based on experience. For small values of α, one or more additional modes comprise ~~of~~ noise. For large values of α, the essential parts of the signal are shared by at least two distinct modes, and their center frequencies overlap, resulting in mode duplication. In addition, when the value of K is set too large, tampering features impede the accuracy of signal filtering, and essential intrinsic mode components are missed when the value of K is set too small. ~~Also~~Additionally, the ~~computation~~computational load can also be large due to the size of the data and a large mode number. ~~So~~Therefore, it is necessary to optimize those values to achieve satisfactory performance.

In the existing optimization techniques~~ologies~~, many sequential search techniques are based on greedy methods. They are~~It is~~ not suitable for global optimality but acceptable for local optimality. For instance, orderly searches consist of forward and backward selection. However, orderly forward and backward search techniques are not only more computationally expensive but also cannot perform undo processes, such as deleting or inserting features. In recent years, a novel emetic ~~genetic algorithm~~GA method for solving the traveling salesman problem was proposed in [1]. An application of GA and fuzzy goal programming to solve congestion management ~~problem~~problems was proposed in [22]. The GA technique is based on evolutionary theory and the random search method. In this case, randomness is added to the search process to avoid local ~~optimum~~optima. GA is reliable and widely used in the~~ area of~~ optimization of artificial neural network parameters or signal processing algorithm parameters [28,37]. Therefore, GA is introduced in this work to obtain the optimal values of the VMD parameters. For the optimization of signal processing parameters, the entropy concept is applied to the GA-VMD algorithm. In theory, a smaller entropy value leads to stronger properties and a clear signal distribution. ~~Minimum~~The minimum envelope spectrum entropy value (MESEV) is proposed as the fitness function of the optimization and is obtained by the following steps:

(i) The Hilbert transform of an IMF signal, which is further described as a time series $\{u_k(t)\}$, can be expressed by

$$h_k(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u_k(t)}{t-\tau} d\tau \quad (6)$$

where t=1,2,...,N, and N is the length of the signal.

(ii) The envelope of the signal $u_k(t)$ is:

$$E_k(t) = \sqrt{u_k^2(t) + h_k^2(t)} \quad (7)$$

(iii) T~~Normalization to t~~he envelope $E(t)$ is normalized as follows:

$$N_k(t) = \frac{E_k(t)}{\sum_{t=1}^{N} E_k(t)} \tag{8}$$

(iv) The envelope spectrum entropy value after normalization is:

$$V_k = -\sum_{t=1}^{N} N_k(t) \ln N_k(t) \tag{9}$$

(v) The ~~minimum envelope spectrum entropy value~~MESEV is:

$$\langle a, K \rangle = \arg\min\{V_k\} \tag{10}$$

The proposed GA-VMD method is summarized in Fig. 2. The initial ranges for parameters *a* and *K* are assigned according to the actual situation at the beginning of the process. Then, ~~it~~ GA-VMD initializes the population of GA and calculates the MESEV of each IMF. The operators in GA are compared to determine whether the current MESEV is the minimum. If not, the population is updated by new individuals until ~~reaching~~ the minimum is reached. MESEV is used as a fitness function~~,~~ so that the iteration is stopped when the minimum MESEV ~~is converged to~~converges to a stable constant or ~~it~~ reaches the preset number of iterations. The values of *a* and *K* at~~under~~ the minimum MESEV are the optimal values.

Sample Entropy IMF selection methods ~~which~~that are commonly used in VMD are presented in this work to select and reconstruct important IMFs. Sample entropy is investigated to determine sensitive subcomponents.

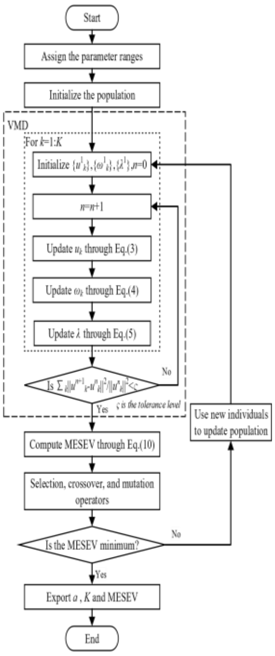


Fig. 2~~Fig.2~~. Flowchart of the GA-VMD method

Even though a higher energy ratio can reflect ~~the~~ fault-related information, faults usually appear ~~in~~at a low energy ratio. Noise always exists in raw signals and ~~that~~ may cause incorrect ~~IMFs~~IMF selections. By d~~D~~efining $N-m+1$ templates, each of size m, which are composed as $F^m(t)=[f(t), f(t+1),..., f(t+m-1)]$;~~, as well as~~ $U_k^m(t)=[u_k(t),u_k(t+1),...,u_k(t+m-1)]$~~;~~, and $t=1,...,N-m+1$, the distance, $d[F^m(t),U_k^m(t)]$, between $F^m(t)$ and $U_k^m(t)$ is computed as $d[F^m(t),U_k^m(t)]=\max|f(t+j)-u_k(t+j)|$, $j=0,...,m-1$. The sample entropy (SampEn) [24] is different from the energy-based method, which is expressed as:

$$\mathrm{SampEn}_k = \ln\left|\frac{\sum_{j=1}^{N-m+1} D_k^m(j)}{N-m+1} - \frac{\sum_{j=1}^{N-(m+1)+1} D_k^{m+1}(j)}{N-(m+1)+1}\right| \tag{11}$$

where $D_k^m(j) = \frac{N_k^m(j)}{N-m+1}$ is the probability that $U_k^m(t)$ matches $F^m(t)$, and $N_k^m(j)$ is defined as the number of template ~~matching~~matches, *i.e.*, the number of $d[F^m(t),U_k^m(t)]<r$. In [23], Pincus suggested that the value of the threshold r should be selected between 0.1 and 0.25 and multiplied by the standard deviation of the raw

signal and that *m* should be equal to 1 or 2. The IMFs with values higher than a preset threshold are chosen as the sensitive IMFs to reconstruct the ~~denoise~~denoised signal.

Remark 1. The above selection algorithms are used to determine the sensitive subcomponents from all IMFs, and the sensitive IMFs ~~could~~can reflect the knock features. The main pure signal is then reconstructed from the selected IMFs;~~,~~ *i.e.,* $\hat{f}(t)=\sum_{\mathcal{P}=1}^{\mathcal{K}} u_{\mathcal{P}}(t)$, where $u_{\mathcal{P}}(t)$ is the $\mathcal{P}$th sensitive ~~IMFs~~IMF decomposed by VMD, and K is the number of~~ the~~ sensitive IMFs.

3.3 Feature Extraction

In this section, a brief description of the three main feature sets used in the proposed multiple feature learning system is given.

Time Domain Statistical Analysis Traditionally, machinery signals were~~are~~ usually extracted by ~~time domain statistical analysis (~~TDSA~~)~~ [29]. These statistical features describe the characteristics of a signal by a direct calculation with simple computations. ~~The features~~ Features such as the standard deviation, root ~~-~~mean ~~-~~square, peak, skewness, kurtosis, crest factor, shape factor and impulse factor are employed in this work.

Alpha Stable Distribution ~~Alpha~~~~The alpha stable distribution alpha stable distribution~~ ~~(~~ASD~~)~~ is suitable for describing random signals ~~having~~ that have ~~a~~ highly non-Gaussian distributions and heavy tails [33]. In ASD, the probability density function (PDF), which is utilized for describing the statistical characteristics of data, can be determined by the four parameters α, β, γ and δ. These parameters are usually expressed by their characteristic functions,

$$\phi(\mathrm{t})=\exp\left(\mathrm{j}\delta\mathrm{t}-\gamma|\mathrm{t}|^{\alpha}\left[1+\mathrm{j}\beta\,\mathrm{sign}(\mathrm{t})\theta(\mathrm{t},\alpha)\right]\right) \tag{12}$$

where $\theta(\mathrm{t},\alpha)=\begin{cases}\tan\left(\frac{\pi\alpha}{2}\right) & \alpha\neq 1\\ \frac{2}{\pi}\log|\mathrm{t}| & \alpha=1\end{cases}$. In this work,~~ the~~ four parameters (α, β, γ and δ) are used to describe the different characteristics as ~~the~~ features for ~~the~~ further classification.

Multif~~-F~~ractal Detrended Analysis ~~Fluctuation Analysis Detrended Multifractal detrended fluctuation analysis (~~DFA~~)~~ is a fractal scaling method for perceiving long-range ~~correlation~~correlations in noisy and nonstationary time sequences. However, DFA is a ~~mono-fractality~~monofractality method and is barely able to deal with ~~multi-fractality~~multifractality nonlinear time series in dynamical ~~mechanism~~mechanisms. Therefore, ~~multi-fractal~~multifractal ~~detrended fluctuation analysis (~~MFDFA~~)~~ was proposed for ~~the~~ multifractality ~~non-stationary~~nonstationary time series analysis by extending the theory of DFA [18]. MFDFA has been verified in revealing the dynamic behavior hidden in ~~multi-scale~~multiscale nonstationary signals~~. which~~ and is described as follows.

The processed bounded time series $\{\hat{f}(1),\ldots,\hat{f}(t)\}$ is converted into an unbounded time series $\{\mathcal{F}(1),\cdots,\mathcal{F}(t)\}$ by a cumulative sum as follows:

$$\mathcal{F}(t)=\sum_{i=1}^{t}(\hat{f}(i)-\bar{f}(t)) \tag{13}$$

where $\bar{f}(t)$ is the mean of the time series~~-~~ $\{\hat{f}(1),\cdots,\hat{f}(t)\}$. Then, F (t) is divided into N_p ~~non-overlapping~~nonoverlapping segments with equivalent lengths p, where N_p≡int(N/p). If N cannot be divided by p, the remaining part of the profile may be ~~left-off.~~truncated. To retain with this unused part, the same process is implemented from the opposite end, and $2N_p$ segments are derived. For segment l=1,...,N_p, the least ~~-~~square of $F^2(p,l)$ is calculated as~~,~~

$$F^2(p,l)=\frac{1}{p}\sum_{i=1}^{p}\left(\mathcal{F}((l-1)p+i)-f_l(i)\right)^2 \tag{14}$$

For segment l=Np+1,...,$2N_p$

$$F^2(p,l)=\frac{1}{p}\sum_{i=1}^{p}\left(\mathcal{F}(N-(l-N_p)p+i)-f_l(i)\right)^2 \tag{15}$$

where $f_l(i)$ is a fitting polynomial in the lth segment. Different ~~order~~orders of the polynomial ~~results~~result ~~in~~ are obtained by different eliminating trends from the profile. The qth order fluctuation function can be obtained by the average over all segments

$$F_q(p)=\left(\frac{1}{2N_p}\sum_{l=1}^{2N_p}\left(F^2(l,p)\right)^{q/2}\right)^{\frac{1}{q}} \tag{16}$$

where q is any real value except zero. Using different time scales of p, the scaling behavior of the fluctuation functions can be determined by analyzing the logarithmic relationship of $F_q(p)$ versus p for each q,

$$F_q(p)\propto p^{H(q)} \tag{17}$$

The~~Setting a~~ relationship between the generalized Hurst exponent $H(q)$ and the scaling exponent $\tau(q)$~~,~~ ~~it~~ is as follows:~~yields~~

$$\tau(q)=qH(q)-1 \tag{18}$$

The singularity exponent h_q and the multifractal singularity spectrum D_q are selected as the features and expressed as~~,~~

$$h_q=\tau'(q)=H(q)+qH'(q) \tag{19}$$

$$D_q=qh_q-\tau(q)=q[h_q-H(q)]+1 \tag{20}$$

where $H^0(q)$ represents the derivative of $H(q)$ with respect to q. The Holder exponent h_q characterizes the strength of the singularity, and D_q represents the Hausdorff dimension of the fractal subset with the exponent h_q, which are utilized to describe the different characteristics.

批注 **[Ed6]:** Please ensure that the intended meaning has been maintained in this edit.

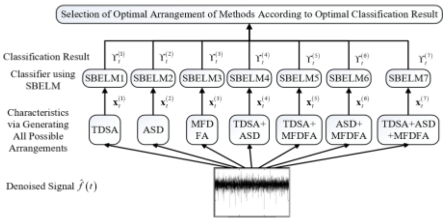


Fig. 3~~Fig.3~~. Multiple feature learning process

Remark 2. The three feature extractors describe the features from three aspects~~,~~ and have ~~multi-~~multiple forms of arrangements and compositions. Time domain features have been proven to be effective for degradation monitoring and failure prognostics in the existing ~~literatures~~literature. MFDFA is able to characterize the internal dynamics mechanism of fault ~~signal~~signals and to detect slight changes in complex ~~environment~~environments. The widely used ASD method has good robustness in the modeling of pulse shape in non-Gauss signals.

Remark 3. The above feature extraction techniques, including the TDSA, MFDFA and ASD methods, and their possible arrangements (*i.e.*, combinations), as shown in Fig. 3, are tested to describe the distinguishable characteristics of the denoised signals~~, respectively~~. The optimal arrangement for finalizing the design of the feature extraction approach, as shown in Fig. 4, is determined according to the optimal classification results obtained through the SBELM classifiers, which are~~is~~ described in the following section.

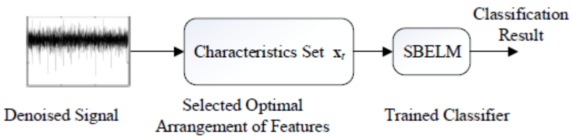


Fig. 4~~Fig.4~~. Final knock detection system

3.4 Sparse Bayesian Extreme Learning Machine for Engine Knock Detection

The ~~sparse Bayesian extreme learning machine (~~SBELM~~)~~ classifier is trained on~~by~~ data (x,T), which contain~~is achieved from~~ the above characteristics of any one arrangement and the known knock label~~, respectively~~. It is well-known that ~~the~~ neural network methods have been used successfully for fault diagnoses, ~~also~~and recently, a~~the~~ family of ~~ELM~~ELMs has been developed for training ~~a~~an SLFN with fast learning speeds and ~~high~~ good generation performance. However, the execution time of ELM is quite unstable and~~,~~ depends~~ing~~ on the number of hidden neurons (network size). Although a kernel-based ELM (KELM) ~~has been proposed~~ that does not require hidden neurons and tends to provide better accuracy than basic ELM has been proposed, it suffers from

~~the issues of a~~ large~~ size of~~ model size issues when the size of the training dataset ~~increases~~is large. Before the development of ELM, RVM was also available. RVM can train the kernel machine on~~for~~ a dataset and automatically prune the irrelevant basis elements to gain sparsity. ~~For~~To reduce the ~~reduction of~~ sensitivity of the number of hidden neurons in conventional ELM, SBELM was proposed~~ combined~~, and it combines the advantages of the low computational load of~~from~~ ELM and the small weight ~~together with~~and good prediction posterior probability of~~from~~ RVM. Reference [17] showed that when the number of hidden nodes is over 50, the classification accuracy could ~~keep~~remain stable. This feature makes it more suitable as a large-scale fault classifier. The SBELM algorithm can be explained as follows.

The output weight of SBELM is learned by the Bayesian method instead of using the Moore-Penrose generalized inverse of the matrix [2]. The hidden layer output H=[h_1~~,~~...,h_t~~,~~...,h_N]^T becomes the input of SBELM, where $h_t \in R^L$ is the hidden feature mapping with respect to input $x_t \in R^L$, L is the number of characteristics of the optimal arrangement, and N is the number of classifier ~~output~~outputs. Each training sample x_t from the extracted features can be treated as an independent Bernoulli case.

Using iterative reweighted least squares to find the Laplace mode Ŵ is efficient; hence,~~, so that~~ the gradient ∇E and Hessian matrix ϕ ~~are necessary to~~must be computed:

$$\nabla E = \nabla_W \ln\{P(\mathrm{T}|\mathrm{W},\mathrm{H})P(\mathrm{W}|\boldsymbol{\alpha})\} = \mathrm{H}^T(\mathrm{T}-\mathrm{Y}) - \mathrm{AW} \tag{21}$$

$$\phi = \nabla_W\nabla_W \ln\{P(\mathrm{T}|\mathrm{W},\mathrm{H})P(\mathrm{W}|\boldsymbol{\alpha})\} = -(\mathrm{H}^T\mathrm{BH}+\mathrm{A}) \tag{22}$$

where W=(w_1~~,~~...,w_m~~,~~...,w_L)^T is the hidden layer matrix. $\mathrm{T} = (\mathcal{T}_1, \ldots, \mathcal{T}_t \ldots, \mathcal{T}_N)^T$, $\mathcal{T}_i \in \{0,1\}$ is a target output vector. α=[α_1~~,~~...,α_L]^T is the independent prior in relation to~~ship with~~ each w_m, and some values of w_m are regulated to zero by adaptive rectangular decomposition~~the~~ (ARD) to select important hidden neurons. Y=(y_1,...,y_N)^T~~ with~~, where y_t=σ(h_t,w_t);~~,~~ A=diag(α); and B is a diagonal matrix, where~~ with~~ b_t=y_t(1-y_t). Subsequently, Ŵ can be obtained by

$$W_{new} = W_{old} - \phi^{-1}\nabla E = (\mathrm{H}^T\mathrm{BH}+\mathrm{A})^{-1}\mathrm{H}^T\mathrm{B}\hat{\mathrm{T}} \tag{23}$$

where $\hat{\mathrm{T}} = \mathrm{HW} + \mathrm{B}^{-1}(\mathrm{T}-\mathrm{Y})$. The center W and covariance matrix Σ of the Gaussian distribution are

$$\Sigma = (\mathrm{H}^T\mathrm{BH}+A)^{-1} \quad \text{and} \quad \hat{\mathrm{W}} = \Sigma\mathrm{H}^T\mathrm{B}\hat{\mathrm{T}} \tag{24}$$

As a result, ln{P(T|W,H)P(W|α)}∝N(Ŵ,Σ) is formed, and the log marginal likelihood L(α)=ln P(T|α,H) can be computed by setting ~~the~~ L(α) to zero, as ~~the~~ follows~~ing expression~~:

$$\frac{\partial L(\boldsymbol{\alpha})}{\partial \alpha_m} = \frac{1}{2\alpha_m} - \frac{1}{2}\Sigma_{mm} - \frac{1}{2}\hat{w}_m^2 = 0 \rightarrow \alpha_m^{new} = \frac{1-\alpha_m\Sigma_{mm}}{\hat{w}_m^2} \tag{25}$$

By setting the initial values of w_m and α_m, $\hat{W}$ and Σ are updated by Eq. (24), and the values of α_m are updated by substituting α_m and Σ into Eq. (25). The marginal likelihood function is iterated to the maximum value until the convergence criterion is met.

In summary, the whole learning procedure of the fault diagnosis scheme is given below. Given the knock label $\mathcal{T}_t$ and the training denoised signal $\hat{f}(t)$, the training procedure is shown as follows.

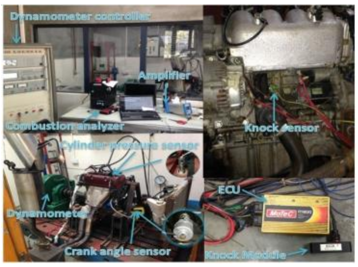


Fig. 5~~Fig.5~~. Test rig

Training procedure

(i) Extract the characteristic data $\mathbf{x}_t^{(r)}$ via generating all possible arrangements of three feature extraction methods from the denoised training signal $\hat{f}(t)$, $r = 1, \ldots, 7$

(ii) For each arrangement,

Initialization: randomly generate input weights and calculate the output of hidden layer H, $\mathrm{W} = \mathbf{0}$, $\boldsymbol{\alpha} = 10^{-5}\mathbf{1}$

Step 1: Estimation of output weights W

(a) Set the initial value $\Sigma = \mathbf{0}$, and define an intermediate variable $g = \mathbf{0}$

(b) Sequentially calculate the mapping of every input $\mathbf{x}_t^{(r)}$ to h_t with random ELM hidden weights

For $t = 1 : N$

$\epsilon = \epsilon + y_t(1 - y_t)h_t^T h_t$

$g = g + (-1)(\mathcal{T}_t - y_t)h_t^T$

End for

(c) $\Sigma = (\epsilon + \mathrm{diag}(\boldsymbol{\alpha}))^{-1}$, $\nabla E = g + \mathrm{diag}(\boldsymbol{\alpha})W$

(d) Find step size λ with line search method, $\mathrm{W} = W - \lambda\Sigma^{-1}\nabla E$

(e) If $norm(\nabla E)$ is under a predefined gradient tolerance, then go to Step 2. Otherwise, go to Step 1.

Step 2: Estimation of hyperparameter $\boldsymbol{\alpha}$.

(f) For every α_m

$\alpha_m = (1 - \alpha_m \Sigma_{mm}^{-1})/w_k^2$

End for

Step 3: Pruning nodes

(g) If α_m >predefined maximum

prune α_m, w_m, $H(:, m)$, $L = L - 1$

End if

(h) If the absolute difference between two successive logarithm values of α_m is lower than given tolerance, then stop. Otherwise, repeat **Step 1 to Step 3.**

(iii) Calculate the classifier results of each arrangement, and select the optimal arrangement.

Testing procedure

For each denoised signal $\hat{f}(t)$,

(i) Extract the characteristic data $\mathbf{x}_i$ via selected optimal arrangement from the de-noised signal $\hat{f}(t)$.
(ii) Calculate the output of the related classifier, whose parameters are inherited from training procedure.

4 Experiment and Evaluation

4.1 Experimental Setup

~~In order to~~ To test and train the proposed framework, a test rig is designed to collect the raw engine data and~~ it~~ is presented ~~as~~ below.

A Honda K20A Type-R engine, which is a four-stroke, four-cylinder spark-ignition engine, is utilized as the test rig, as shown in Fig. 5. The research octane number of the fuel is 98, which was~~is~~ purchased from a regular gas station. The experimental setup ~~as~~ shown in Fig. 6 can be divided into three main sections. The first section contains~~is~~ the ECU~~electronic control unit~~, the engine and relative peripheral sensors, where the raw data ~~is~~are collected via a knock sensor. The second section contains~~is~~ the dynamometer and its control system for varying the loading condition of the engine. The third section contains~~is~~ the combustion analyzer with an in-cylinder pressure sensor, which is used to detect whether knock exists in the experiment. The data collected by the in-cylinder pressure can validate the result of the proposed system. The main components are as follows:

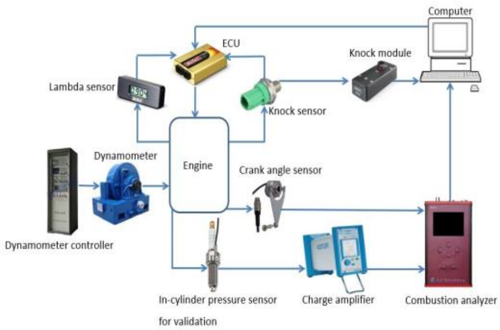


Fig. 6~~Fig.6~~. Test rig setup

Electronic Control Unit A MoTeC M800 programmable ECU ~~can~~ controls the engine by monitoring sensor signals and adjusting the outputs based on the look-up tables. The ECU can control the spark timing, fuel injection time and engine temperature~~, etc~~. In this work, the injection time and ignition timing are important for ECU control. During the experiment, the injection time and ignition timing at different engine speeds and loads can be adjusted through the fuel map and ignition map in the ECU, respectively. The fuel map mainly controls the air-fuel ratio or air~~-~~ ratio. To measure the air-fuel ratio/air-ratio, a lambda sensor/oxygen sensor is installed in the exhaust pipe and used for measurement.

Dynamometer and Control System A DW160 eddy-current dynamometer is used to apply the engine load and control the engine throttle for simulating different driving conditions. The dynamometer is coupled to the test engine.

Combustion Analyzer A *MA3001* combustion analyzer, which ~~is~~was produced by PowerMAC Co., Ltd., is used ~~for analyzing~~to analyze the in-cylinder pressure and corresponding crank angle. The analyzer consists of two parts: (i) The c~~C~~rank angle sensor, which is mounted on the engine crankshaft terminal to measure the engine crank angle in the engine cycle. The sensor is used to convert the rotational speed and phase position of the crankshaft into a digital angle signal, which helps monitor the pressure wave for knock detection. (ii) A piezoelectric in-cylinder pressure sensor is employed to measure the in-cylinder combustion pressure for validation. The signal from the cylinder pressure sensor is then amplified by a charge amplifier. The crank angle signal and the amplified in-cylinder pressure signal are sent to the analyzer for pressure wave analysis. Before starting the experiment, the devices ~~have~~had to be calibrated. The calibrated range and sensitivity charge of the amplifier ~~is~~are set to ~~150*bar*~~*150 bar* and −10.~~22*pC/bar* in order~~ 22 pC/bar to match the in-cylinder pressure sensor. The mode of the amplifier is set to 0–1~~0 10*V*~~*10 V* according to the specification of the combustion analyzer. The voltage-pressure conversion coefficient of the combustion analyzer is set to ~~be~~ 15, depending on the amplifier and the test engine torque. It is worth noting that the top dead center position needs to be calibrated when the crank angle sensor is installed on the test engine.

Data Collection and Analysis A software called *GoldWave* is installed on a computer to record the engine signals from the knock sensor. The signal is then passed to MATLAB to conduct signal filtering, feature extraction and classification.

4.2 Operating C~~c~~onditions for ~~experiment~~E~~e~~xperimental D~~d~~ata C~~c~~ollection

~~In order to~~ To verify the proposed scheme, real engine data ~~is~~are recorded and analyzed. Since the fuel used in the experiment has a high-octane ~~high octane~~ number, engine knock ~~is~~does not ~~easy to~~ easily occur. ~~In order to~~ To produce ~~a~~ knock ~~condition~~conditions under different driving conditions without damaging the engine in the laboratory~~lab~~, the engine is operated under two working conditions: i) low speed with high load ~~condition,~~ conditions and ii) high speed with low load ~~condition~~conditions. The engine load is provided by the dynamometer by applying ~~an~~ opposite torque to the engine. The ignition timing is advanced gradually. The initial engine temperature before knocking ~~is hold~~is held at ~~85 °C~~85°C ± ~~5 °C~~5°C. The engine load, speed and air-fuel ratio are changed within a certain range. The combustion analyzer records the pressure wave pattern to determine the presence of engine knock so that the training and test data can be obtained. A total of 1800 sets of data are recorded according to different driving conditions, as shown in Table 1.

Table 1. Experimental data

Operation condition				Number of samples	Objective
Speed (rmp)	Load (Nm)	Air-fuel	Ignition Timing (°BTDC)		
1000±300	60 ± 5	1±0.5	10°±2.5 to 45°±2.5	320	Simulate a low speed and high load driving condition
2000±300	12 ± 5	1±0.5	10°±2.5 to 45°±2.5	990	Simulate a high speed and low load driving condition
3000±300	12 ± 5	1±0.5	10°±2.5 to 45°±2.5	490	Simulate a high speed and low load driving condition

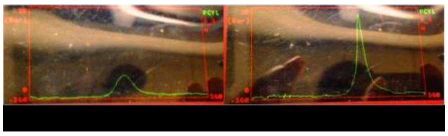


Fig. 7~~Fig.7~~. Pressure versus crank angle ~~at~~under (~~Left~~left) nonk~~non-k~~nock and (~~Right~~right) knock conditions

At the beginning of the experiment, knock does not occur easily at idle speeds due to the high anti-knock quality of the fuel, even ~~if~~ when the ignition timing is substantially advanced ~~very much~~ and the air-fuel ratio is enriched. Under this condition, the cylinder pressure wave pattern in the combustion analyzer is still smooth, as shown ~~in~~on the left~~-hand~~-hand side of Fig. 7. When the ignition ~~advance~~advances and the engine load ~~are kept increasing~~continues to increase, the shape of the pressure wave sharply increases. When~~Until~~ the ignition timing and engine load are increased to a certain range, an obviously high and sharp pressure wave appears, indicating the existence of knock, as shown ~~in~~on the right~~-hand~~-hand side of Fig. 7. Therefore, it is not easy to generate a knock at a low engine speed with~~under~~ a high-octane fuel unless the engine load is high. Certainly, ~~engine operates~~engines operating at a high engine speed under a high-octane ~~high octane~~ fuel can generate ~~a~~ knock easily under a low engine load. It is noteworthy that the combustion analyzer and in-cylinder pressure signal are not suitable for in-use vehicles due to their high costs, so they are used only for validation and labeling ~~only~~. The actual knock detection signal is the engine vibration signal captured by~~from~~ the knock sensor.

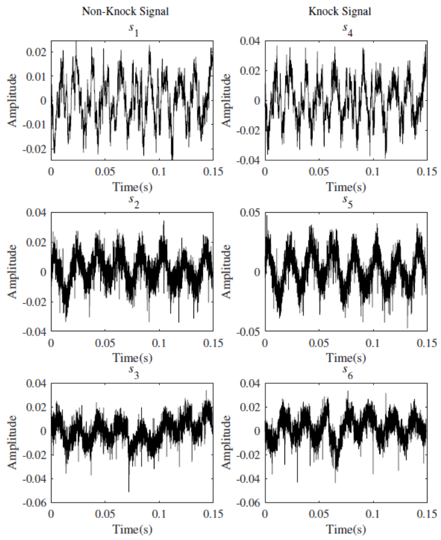


Fig. 8~~Fig.8~~. Time domain engine vibration signals

Table 2. Experimental setup of the sample vibration signals

Vibration Signal		Speed (rpm)	Load (Nm)	Air-fuel ratio	Ignition timing (°BTDC)
Non-knock	s_1	1000	58.1	0.9	20
	s_2	2000	7	0.9	20
	s_3	3000	9	0.9	20
Knock	s_4	1000	58.1	0.9	40
	s_5	2000	12	0.9	40
	s_6	3000	15	0.7	42

The vibration signal collected by the knock sensor converts the shock of cylinder pressure into an electronic signal. For each driving condition, the raw signals are recorded for 0.15 ~~second~~seconds with a sampling rate of ~~48000Hz~~48000 Hz. Therefore, each sample ~~contain~~contains a time series with 7200 sampling points. ~~6~~Six randomly selected vibration signals from the 1800 sets of data shown in Table 2 are illustrated in Fig. 8, ~~with~~ where half of the signals are nonk~~non-k~~nock labeled signals and half are knock labeled signals. They are used as training ~~dataset~~datasets to train the classifiers. It can be observed from Fig. 8 that the nonk~~non-k~~nock signals (s_1, s_2, s_3) are very~~quite~~ difficult to manually distinguish ~~the difference manually~~ from the knock signals (s_4, s_5, s_6). Therefore, the proposed framework is applied to remove ~~the noises~~ noise from ~~the~~

vibration signals and ~~detects~~detect knock. The experimental data and program code in ~~Matlab~~MATLAB are available at https://github.com/wangdai11/EKDS.

4.3 Results and Evaluation

Signal Filtering Signal filtering is the first step of the proposed framework, and~~to~~ it reduces noise from the raw vibration signals. VMD converts the raw signals into a series of IMFs. Sample entropy is employed ~~into~~in the proposed signal processing methods to remove the insensitive IMFs. For comparison, signal s_6, is used~~as~~ an example in this section~~, is utilized~~ to evaluate the filtering ability of the proposed GA-VMD.

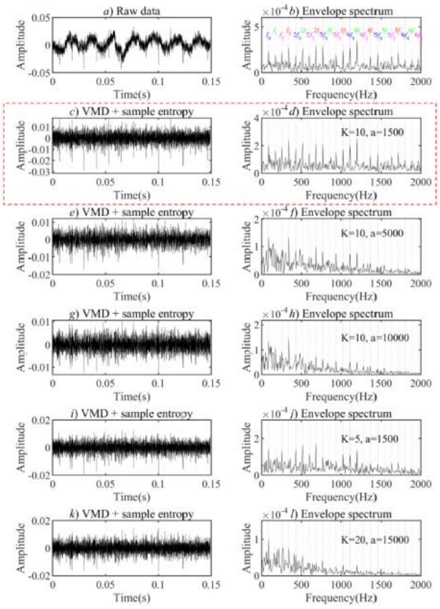
IMFs of VMD ~~depends~~depend on the adjustable ~~parameter~~parameters α and K, which are inaccurate when the parameters are set inappropriately. Therefore, GA is proposed to obtain the appropriate values for α and K. The parameters of GA are set as follows: population size=50, number of generations=200, mutation rate=0.01, mutation percentage of the~~on~~ population=0.2, and crossover percentage of the~~on~~ population=0.8. The input ranges of α and K are set to [100,10000] and [2,20], respectively. ~~Taking the average of each optimal value of α and K a~~After 50 runs of GA, the average values are α=1463 and K=9.9, respectively. Therefore, α and K are set to 1500 and 10.

Fig. 9 illustrates~~is~~ an example that shows the influence of~~ setting~~ different values of α and K on signal filtering. When α is set too large or when K is set inappropriately, some knock resonant frequencies (Fig. 9*f*, 9h~~9*h*~~, 9*j*, and 9l~~9*l*~~) cannot ~~display~~be clearly ~~as~~displayed compared with Fig. 9*b*. Choosing sample entropy as the IMF selection method due to the best noise reduction ability, Fig. 9*c* and Fig. 9*d* show the GA-VMD results. Fig. 9*c*, Fig. 9*d* and Appendix A show that only~~ the~~ GA-VMD can clearly reflect all the resonant frequencies~~ clearly~~.

The results of using VMD and different IMF selection methods for signal s_6 are shown in Fig. 10 and Table 3. Each method takes the threshold T to select the appropriate IMFs for signal reconstruction, where $T = \frac{\sum_{n=1}^{K} IMF_K}{K}$ and K is the total number of IMFs. Th~~os~~e IMFs with values higher than the threshold are chosen and highlighted in red in Table 3. Th~~os~~e selected IMFs are reconstructed into a denoising signal, and the envelope spectrum of the filtered signals is used to identify the knock resonant frequency. Fig. 11 shows the envelope spectrum of the GA-VMD noise reduction under different IMF selection~~ed~~ methods. Fig. 11*h* and Appendix B show that only the sample entropy can reflect the knock resonant frequencies, as shown in Fig. 11*e*~~ clearly~~. This further indicates that the sample entropy approach has ~~a ~~good noise reduction and signal reconstruction abilities.

批注 **[Ed7]:** Please ensure that the intended meaning has been maintained in this edit.

带格式的: 字体: 加粗



~~Fig. 9~~Fig. 9. Noise reduction ability under different values of *a* and *K*

Table 3. Results of GA-VMD with different IMF selection methods

s_6	Correlation Coefficient	Energy Ratio	Sample Entropy
IMF_1	0.7790	0.5632	0.0645
IMF_2	0.3963	0.0744	0.5218
IMF_3	0.3051	0.0356	0.6086
IMF_4	0.2535	0.0236	0.5841
IMF_5	0.2338	0.0215	0.5775
IMF_6	0.2086	0.0162	0.5748
IMF_7	0.1954	0.0147	0.5362
IMF_8	0.1875	0.0128	0.5934
IMF_9	0.1934	0.019	0.5911
IMF_{10}	0.1442	0.0084	0.5979
T	0.2897	0.0790	0.5260

Feature Extraction Feature extraction, a pretreatment for machine learning ~~method~~methods, is the second step of the proposed knock detection method. The ~~application~~applications of TDSA, ASD and MFDFA are used for extracting cognizable features from the filtered signals. Each extracted feature can compress a ~~huge~~ large number of time series data into specific numbers. These specific numbers representing meaningful features are then used to establish a classification model for knock detection.

Table 4 shows the TDSA features of 24 randomly selected ~~24~~ engine vibration signals under different conditions, including mean y_{mean}, standard deviation y_{std}, root ~~-~~mean square y_{rms}, peak y_{peak}, skewness y_{skew}, kurtosis y_{kurt}, crest factors y_{crf} and y_{clf}~~-~~, shape factor y_{sf} and impulse factor y_{if}, which are created under different ignition timing and loading conditions. In Table 4, the sample signals A_1 to A_8 are at ~~of~~ 1000 rpm, B_1 to B_8 are ~~of~~at 2000 rpm and C_1 to C_8 are at ~~of~~ 3000 rpm. These statistical features can be used to separate knock data from ~~the~~ nonk~~non-k~~nock data. Therefore, these statistical features are kept for the inputs of the classifiers.

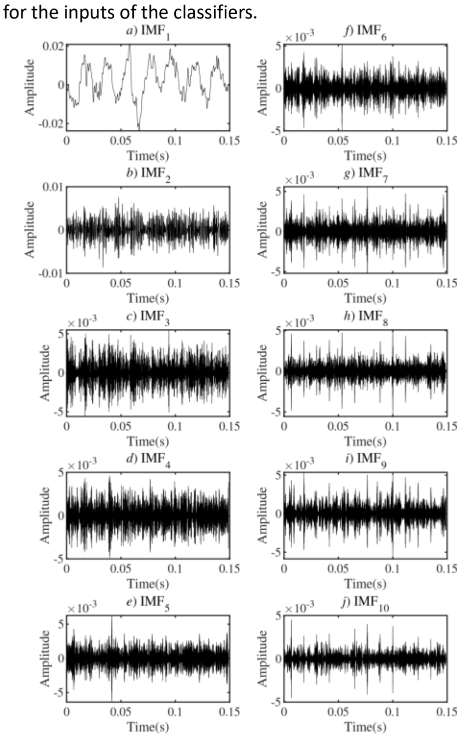


Fig. 1~~Fig. 1~~0. IMFs obtained based on GA-VMD

The ASD algorithm is a feature extraction method that emphasizes the characteristic parameters α, β, γ, and δ. The values of these parameters are self-generated by the wave patterns of the signal. The ASD characteristic parameters and the magnitudes of the PDF are different under knock ~~or~~ and nonk~~non-k~~nock ~~condition~~conditions, as shown in Fig. 12. Therefore, the ~~parameter~~parameters α, β, γ, δ and *h* are selected as the inputs of the classifiers. Table 6 shows the five ASD parameters of the same 24 vibration samples (A_1 to A_8, B_1 to B_8 and C_1 to C_8) in Table 4.

Fig. 13 depicts that the knock data mainly lay between the large values of γ and α, but the nonk~~non-k~~nock data are dispersive. Most of the nonk~~non-k~~nock data have higher values of *h* and α than the knock data. In this case, most ~~of~~ knock data can be separated from the nonk~~non-k~~nock data with~~under~~ this method.

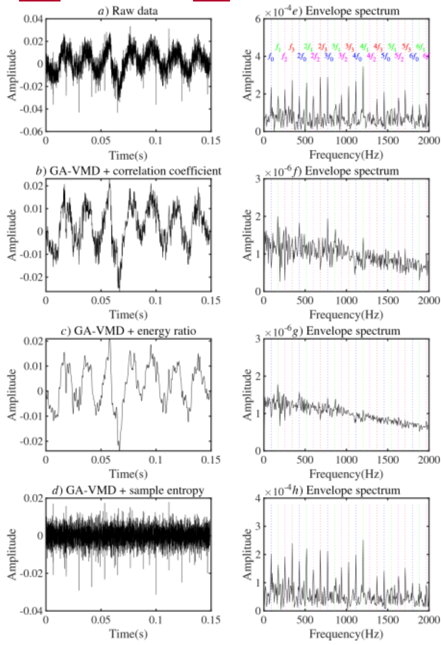


Fig. 1~~Fig.1~~1. Envelope spectrum of GA-VMD under different IMF selection methods

MFDFA is another feature extraction approach ~~which~~that emphasizes the 3 points in the multifractal spectrum: i) ~~The~~the first points of the multifractal curves (h_{qa-},D_{qa-}); ii) ~~The~~the end points of the multifractal curves (h_{qb-},D_{qb-}); and iii) ~~The~~the peaks of the multifractal curves (h_{q0},1). The signal under various working conditions ~~provide~~provides different ~~spectrums~~spectra, as shown in Fig. 14. Table 7 shows the five multifractal parameters (h_{qa-}, D_{qa-}, h_{qb-}, D_{qb-}, and h_{q0}) of the same 24 vibration samples (A_1 to A_8, B_1 to B_8 and C_1 to C_8). The distribution results of the multifractal parameters in Fig.~~ure~~ 15 show that most of the knock data in Table 4 can also be separated from the nonk~~non-k~~nock data under GA-VMD. Therefore, MFDFA is also considered in this work.

Table 4. Example of the TDSA result of GA-VMD+Sample entropy

		y_{mean}	y_{std}	y_{rms}	y_{peak}	y_{skew}	y_{kurt}	y_{crf}	y_{clf}	y_{sf}	y_{if}
Non-Knock	A_1	2.29×10^{-7}	9.87×10^{-3}	0.001	0.010	-0.112	4.508	5.318	8.438	1.315	6.991
	A_2	-4.19×10^{-8}	9.70×10^{-3}	0.001	0.010	0.124	5.063	5.842	9.351	1.322	7.721
	A_3	-4.80×10^{-7}	2.60×10^{-3}	0.003	0.011	0.017	3.000	4.002	5.913	1.253	5.014
	A_4	-3.40×10^{-7}	2.36×10^{-3}	0.002	0.010	0.004	3.046	4.108	6.124	1.258	5.169
	B_1	2.56×10^{-6}	2.42×10^{-3}	0.002	0.020	0.151	5.609	7.186	11.518	1.326	9.527
	B_2	8.60×10^{-7}	4.11×10^{-3}	0.004	0.030	-0.098	4.153	6.326	9.556	1.274	8.060
	B_3	2.94×10^{-6}	2.47×10^{-3}	0.002	0.022	-0.073	5.251	7.651	12.234	1.323	10.124
	B_4	2.01×10^{-6}	3.72×10^{-3}	0.004	0.036	-0.242	4.882	7.204	11.008	1.285	9.257
	C_1	5.28×10^{-7}	2.09×10^{-3}	0.002	0.029	0.103	7.403	10.688	17.640	1.354	14.470
	C_2	9.29×10^{-7}	3.87×10^{-3}	0.004	0.037	-0.273	4.680	7.435	11.426	1.289	9.582
	C_3	-6.60×10^{-6}	4.39×10^{-3}	0.004	0.037	-0.266	4.537	7.492	11.530	1.289	9.656
	C_4	-3.10×10^{-6}	4.39×10^{-3}	0.004	0.043	-0.308	4.789	7.037	10.916	1.296	9.122
Knock	A_5	-5.66×10^{-7}	4.11×10^{-3}	0.004	0.014	-0.095	2.994	5.204	1.253	1.253	4.408
	A_6	6.31×10^{-7}	3.28×10^{-3}	0.003	0.014	0.037	3.099	6.204	1.253	1.253	5.433
	A_7	-2.13×10^{-6}	3.29×10^{-3}	0.003	0.014	-0.068	3.159	6.535	1.263	1.263	5.509
	A_8	-5.12×10^{-7}	4.05×10^{-3}	0.004	0.017	0.054	3.642	6.534	1.278	1.278	5.506
	B_5	8.55×10^{-6}	5.29×10^{-3}	0.005	0.027	-0.053	3.604	7.890	1.279	1.279	6.662
	B_6	4.94×10^{-7}	4.91×10^{-3}	0.005	0.027	-0.041	3.789	8.380	1.281	1.281	7.027
	B_7	-1.52×10^{-6}	5.45×10^{-3}	0.005	0.035	-0.103	3.817	9.849	1.275	1.275	8.298
	B_8	8.95×10^{-6}	6.06×10^{-3}	0.006	0.030	-0.100	3.577	7.410	1.271	1.271	6.246
	C_5	9.29×10^{-7}	3.87×10^{-3}	0.004	0.029	-0.183	4.496	11.467	1.281	1.281	9.662
	C_6	-6.82×10^{-6}	5.21×10^{-3}	0.005	0.033	-0.192	4.196	9.614	1.283	1.283	8.068
	C_7	5.41×10^{-6}	6.16×10^{-3}	0.006	0.030	-0.070	3.528	7.338	1.278	1.278	6.156
	C_8	1.49×10^{-5}	5.33×10^{-3}	0.005	0.029	0.023	3.676	8.363	1.279	1.279	7.013

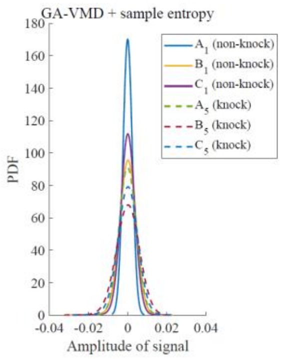


Fig. 1~~Fig.1~~2. PDF spectrum of different signals

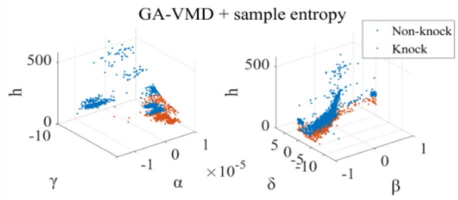


Fig. 1~~Fig.1~~3. ASD parameters

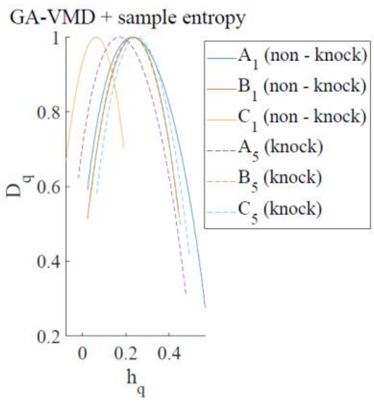


Fig. 1~~Fig.1~~4. Multifractal spectra~~um~~ of different signals

The above three feature extraction ~~method~~methods are feasible and produce different separable features, so they are used in different combinations. In total, 20 features based on the TDSA, ASD and MFDFA methods are obtained, as shown in Table 5. The d~~D~~ifferent characteristics of the knock data and nonk~~non-k~~nock data are then entered ~~to~~into the machine learning methods for building classifiers for diagnosis.

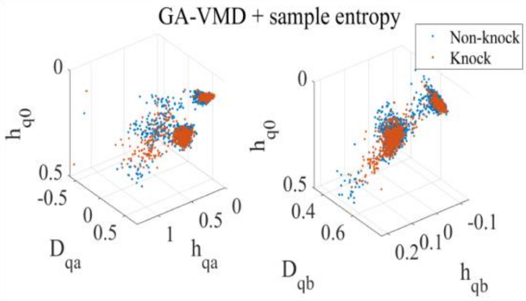


Fig. 1~~Fig.1~~5. MFDFA parameters

Table 5. Extracted features

Methods	Features	Total
TDSA	Mean, standard deviation, root-mean-square, peak, skewness, kurtosis, crest factor, clearance factor, shape factor, impulse factor	10
ASD	$\alpha, \beta, \gamma, \gamma, h$	5
MFDFA	$h_{q_a}, D_{q_a}, h_{q_b}, D_{q_b}, h_{q_0}$	5

Table 6. ASD results with GA-VMD+Sample entropy

		α	β	γ	δ
Non-Knock	A_1	2.000	-1.000	1.90×10^{-3}	-2.86×10^{-5}
	A_2	1.957	-0.112	2.32×10^{-3}	1.60×10^{-5}
	A_3	1.847	0.154	1.02×10^{-3}	4.39×10^{-5}
	A_4	2.000	1.000	2.06×10^{-3}	6.37×10^{-6}
	B_1	1.968	-1.000	2.91×10^{-3}	-1.89×10^{-5}
	B_2	1.951	-0.448	2.95×10^{-3}	5.02×10^{-7}
	B_3	1.805	0.076	1.29×10^{-3}	1.91×10^{-6}
	B_4	1.977	0.609	2.63×10^{-3}	3.54×10^{-5}
	C_1	1.966	-1.000	2.96×10^{-3}	-3.73×10^{-5}
	C_2	1.793	-0.389	1.49×10^{-3}	-4.61×10^{-5}
	C_3	1.828	0.079	1.30×10^{-3}	1.28×10^{-5}
	C_4	1.795	0.529	1.55×10^{-3}	7.55×10^{-5}
Knock	A_5	1.990	-1.000	3.30×10^{-3}	-1.48×10^{-5}
	A_6	1.972	-0.173	3.12×10^{-3}	1.50×10^{-5}
	A_7	1.976	0.068	2.99×10^{-3}	3.95×10^{-5}
	A_8	1.988	1.000	3.33×10^{-3}	7.06×10^{-6}
	B_5	1.964	-1.000	4.05×10^{-3}	4.44×10^{-5}
	B_6	1.927	-0.523	4.08×10^{-3}	2.03×10^{-6}
	B_7	1.943	-0.305	4.52×10^{-3}	7.81×10^{-6}
	B_8	1.976	0.627	4.60×10^{-3}	-9.65×10^{-6}
	C_5	1.952	-1.000	3.56×10^{-3}	-5.84×10^{-5}
	C_6	1.926	-0.295	3.04×10^{-3}	-1.30×10^{-5}
	C_7	1.929	0.013	2.99×10^{-3}	1.93×10^{-5}
	C_8	1.962	0.163	4.08×10^{-3}	2.05×10^{-5}

Classification Classification is the last step of the proposed framework. The extracted features are learned using two other ~~two~~ machine learning algorithms, where ELM and ~~kernel based -~~-based ~~ELM (~~KELM~~)~~ are applied for comparison. To verify the statistical performance of the test results, we use ~~the~~ bootstrapping for the dataset. Bootstrapping is a test or metric that relies on random sampling with replacement. The dataset is separated into two groups, nonk~~non-k~~nock data and knock data, wherein 900 sets are randomly selected as training data and the remaining~~rest of~~ 900 sets are used as test data. The division of the training and test datasets~~data sets~~ is presented in Table 8. The mean results are achieved ~~by repeating~~after 10 repetitions~~times~~ and are shown in Table 9.

Table 7. MFDFA results with GA-VMD+Sample entropy

		h_{q0}	h_{qa}	D_{qa}	h_{qb}	D_{qb}
Non-Knock	A_1	0.098	-0.090	0.599	0.370	0.356
	A_2	0.215	0.012	0.602	0.584	0.162
	A_3	0.317	0.078	0.548	0.692	0.204
	A_4	0.230	0.019	0.590	1.099	-0.626
	B_1	0.050	-0.042	0.793	0.164	0.727
	B_2	0.058	-0.064	0.713	0.238	0.525
	B_3	0.202	0.010	0.593	0.439	0.449
	B_4	0.278	0.048	0.454	0.547	0.347
	C_1	0.048	-0.077	0.702	0.160	0.746
	C_2	0.062	-0.070	0.699	0.235	0.566
	C_3	0.244	0.061	0.579	0.443	0.547
	C_4	0.263	0.035	0.506	0.563	0.271
Knock	A_5	0.119	-0.070	0.600	0.455	0.231
	A_6	0.243	0.035	0.599	0.561	0.308
	A_7	0.340	0.105	0.520	0.693	0.243
	A_8	0.429	0.181	0.517	1.316	-0.537
	B_5	0.049	-0.065	0.735	0.159	0.749
	B_6	0.204	0.015	0.575	0.386	0.601
	B_7	0.237	0.060	0.609	0.468	0.458
	B_8	0.258	0.041	0.539	0.559	0.273
	C_5	0.044	-0.063	0.738	0.151	0.733
	C_6	0.071	-0.067	0.687	0.222	0.641
	C_7	0.241	0.047	0.570	0.447	0.534
	C_8	0.273	0.088	0.607	0.523	0.405

Table 8. Details of the training and testing datasets

Group	Label	Number of training data	Number of test data	Total
1	Non-knock	550	550	1100
2	Knock	350	350	700
	Total	900	900	1800

Table 8 shows that this knock detection problem is a binary classification problem. ~~In order to~~ To select an appropriate classification method, the accuracies of the three machine learning methods are compared. For ELM and SBELM, the number of initial hidden neurons ~~have to~~must be defined. The initial hidden neurons for ELM and SBELM are set to 200. For KELM, the kernel is introduced to the model~~,~~; thus, the regularization parameter and kernel parameter have to be set. The kernel function of KELM is a radial basis function. The regularized parameter and the kernel parameter of KELM are set to ~~be~~ 1.0. The test accuracies are shown in Table 9, and the best accuracy is highlighted in red. Table 9 shows that the average accuracy of SBELM is slightly higher than those of KELM and ELM because the parameters of SBELM ~~is~~are not sensitive to its hyperparameters.

Table 9 reveals that the~~ integrated~~ features of GA-VMD integrated with~~,~~ sample entropy, TDSA, ASD and SBELM ~~show~~ have the best accuracy of 98.27%, which is highlighted in red in ~~this test~~the table. It is noted that ASD and TDSA ~~produce~~ have high classification accuracies~~ in classification~~, whereas MFDFA ~~shows~~ has poor performance. Even though combining MFDFA with other feature extraction methods can improve the overall precision ~~a little bit~~slightly, MFDFA does not contribute too much to the system accuracy. It also appears~~seems~~ that MFDFA is not compatible with GA-VMD because it has~~produces~~ the worst accuracy. In summary, Table 9 shows that the integration of SBELM with GA-VMD, sample entropy, ASD and TDSA is an accurate classification method for automatic knock detection.

Table 9. Accuracies of various combinations of technologies based on the test dataset

Feature extraction	Signal filtering method	ELM	KELM	SBELM
TDSA	Raw data	93.17%	93.71%	93.50%
	EEMD+sample entropy	94.36%	95.12%	95.23%
	GA-VMD+sample entropy	95.66%	97.72%	97.62%
ASD	Raw data	87.43%	91.76%	91.44%
	EEMD+sample entropy	89.49%	91.87%	92.63%
	GA-VMD+sample entropy	92.84%	92.52%	95.88%
MFDFA	Raw data	72.80%	75.40%	74.65%
	EEMD+sample entropy	70.63%	73.23%	73.56%
	GA-VMD+sample entropy	63.59%	64.78%	63.92%
TDSA+ASD	Raw data	92.41%	93.71%	93.50%
	EEMD+sample entropy	94.47%	95.44%	94.58%
	GA-VMD+sample entropy	96.09%	97.72%	98.27%
TDSA+MFDFA	Raw data	92.84%	94.04%	93.72%
	EEMD+sample entropy	94.69%	95.34%	95.12%
	GA-VMD+sample entropy	95.23%	97.39%	96.86%
ASD+MFDFA	Raw data	94.25%	94.36%	93.72%
	EEMD+sample entropy	91.65%	95.44%	94.37%
	GA-VMD+sample entropy	94.04%	95.88%	96.86%
TDSA+ASD +MFDFA	Raw data	93.82%	93.71%	94.37%
	EEMD+sample entropy	93.39%	95.34%	94.26%
	GA-VMD+sample entropy	95.44%	97.18%	97.40%

5 Conclusion

In this paper, a novel intelligence engine knock detection system using a multiple feature~~-based~~ -based SBELM algorithm~~sparse Bayesian extreme learning machine~~ is successfully developed. GA-VMD is used to filter the unavoidable noise~~s~~, in which GA is applied to obtain the optimal parameters to enhance the noise reduction ability. When the original time domain signals are decomposed into a series of IMFs, IMFs with sample entropy higher than the mean ~~is~~are selected as sensitive subcomponents for signal reconstruction. Multiple ~~features~~methods, including the TDSA, MFDFA and ASD methods, are applied together to extract features from the denoised signals. The ~~Extracted~~ features extracted from the reconstructed signals are then classified by SBELM. The e~~E~~xperimental results show that the accuracy of the knock detection system built by SBELM is superior to the accuracies of those built by ELM and KELM. Therefore, the integration of GA-VMD~~,~~ with sample entropy~~, combined with~~ TDSA~~, and~~ ASD, and SBELM is effective for building automatic engine knock detection ~~system~~systems. Although the proposed method is successfully applied to real ~~engine~~engines for engine knock detection, the dataset is recorded from a~~the~~ specific engine model. It will be~~is~~ ~~appealed~~appealing to apply~~ to~~ different engine models to further prove ~~its~~ the reliability of the proposed method in ~~the~~ future work. Moreover, the training and test data for the proposed system can~~would~~ be expanded to cover more engine speeds, engine loads, air-fuel ratios, fuel octane numbers and engine temperatures~~ in order~~ to enhance the system generalization. In our current work, the proposed GA-VMD method has the limitation of eliminating~~ the~~ non-Gaussian noise under ~~the~~ heavy noise disturbances. Non-Gaussian ~~noises~~noise always ~~exist~~exists in ~~the~~ automotive propulsion systems and~~, which~~ usually ~~lead~~leads to inconsistenci~~es~~ies~~y~~ and the divergence of the detection system. Therefore, ~~the~~ future work ~~would~~should consider the noise rejection capacity by using ~~the~~ correntropy to cope with the issue of non-Gaussian ~~noises~~noise.

A Novel Multiple Feature-based Engine Knock Detection System using Sparse Bayesian Extreme Learning Machine

Zhao-Xu Yang, Hai-Jun Rong, Pak Kin Wong ⋆, Plamen Angelov, Chi Man Vong, Chi Wai Chiu, and Zhi-Xin Yang

Abstract.
Background Automotive engine knock is an abnormal combustion phenomenon that affects engine performance and lifetime expectancy, but it is difficult to detect. Collecting engine vibration signals from an engine cylinder block is an effective way to detect engine knock.
Methods This paper proposes an intelligent engine knock detection system based on engine vibration signals. First, filtered signals are obtained by utilizing variational mode decomposition (VMD), which decomposes the original time domain signals into a series of intrinsic mode functions (IMFs). Moreover, the values of the balancing parameter and the number of IMF modes are optimized using genetic algorithm (GA). IMFs with sample entropy higher than the mean are then selected as sensitive subcomponents for signal reconstruction and subsequently removed. A multiple feature learning approach that considers time domain statistical analysis (TDSA), multi-fractal detrended fluctuation analysis (MFDFA) and alpha stable distribution (ASD) simultaneously, is utilized to extract features from the denoised signals. Finally, the extracted features are trained by sparse Bayesian extreme learning machine (SBELM) to overcome the sensitivity of hyperparameters in conventional machine learning algorithms.
Results A test rig is designed to collect the raw engine data. Compared with other technology combinations, the optimal scheme from signal processing to feature classification is obtained, and the classification accuracy of the proposed integrated engine knock detection method can achieve 98.27%.
Conclusions We successfully propose and test a universal intelligence solution for the detection task.

⋆ Corresponding author: Pak Kin Wong ✉fstpkw@um.edu.mo
Zhao-Xu Yang is with State Key Laboratory for Strength and Vibration of Mechanical Structures, Shaanxi Key Laboratory of Environment and Control for Flight Vehicle, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an, China, and State Key Laboratory of Internet of Things for Smart City, Department of Electromechanical Engineering, University of Macau, Macao, China. Hai-Jun Rong is with State Key Laboratory for Strength and Vibration of Mechanical Structures, Shaanxi Key Laboratory of Environment and Control for Flight Vehicle, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an, China. Pak Kin Wong and Chi Wai Chiu are with Department of Electromechanical Engineering, University of Macau, Macao, China. Plamen Angelov is with School of Computing and Communications, Lancaster University, Lancaster, United Kingdom. Chi Man Vong is with Department of Computer and Information Science, University of Macau, Macao, China. Zhi-Xin Yang is with State Key Laboratory of Internet of Things for Smart City, Department of Electromechanical Engineering, University of Macau, Macao, China.

Keywords: Engine Knock Detection, Variational Mode Decomposition, Multiple Feature Learning, Sample Entropy, Sparse Bayesian Extreme Learning Machine.

1 Introduction

In a spark-ignition automotive engine, engine knock is defined as an abnormal combustion phenomenon that is observed as a noise and can even indicate a major engine fault. Engine knock is an essential factor that constrains further improvements in thermal efficiency and causes serious engine damage, such as piston or cylinder breakage. During heavy knock, much extra heat is transferred to the combustion chamber wall, resulting in a rapid rise in the temperature of the piston and cylinder head. The overheating of these parts makes the intensity of the knock continue to increase. This consequential runaway phenomenon may trigger engine failure within a few minutes. Moreover, an excessively high pressure pulse due to heavy knock may occur in the end gas area. The interaction between this high local pressure and high local surface temperature inevitably weakens or corrodes the engine material.

On the premise of the accurate identification of signals associated with engine knock, some preventive measures should be performed, such as delaying the ignition timing. These signals can be monitored and collected by pressure wave amplitude analysis, exhaust gas temperature analysis, heat transfer analysis, *etc.* [40]. However, the high cost of in-cylinder pressure sensors, as well as the decreased lifetime expectancy resulting from high temperatures and high pressure, make the pressure wave amplitude method difficult to apply extensively [11]. Exhaust gas temperature analysis suffers from low precision [19], and heat transfer analysis is difficult to apply in real time [20].

Massive pressure waves [21] occur inside of an ignition chamber, and can emit an audible sound, and the resulting vibrations create the perceptible knock signal. Therefore, engine vibration signals are widely used for engine knock detection, which is a compromised solution for resolving conflicts between measuring precision and cost. When vibrations are detected in the cylinder wall, the knock sensor, which is a piezoelectric crystal placed on an engine cylinder block, creates a low voltage signal that is fed back to the electronic control unit (ECU). Knock can be determined when the resonant frequency is close to or beyond the frequency range of the knock frequency. However, engine vibrations include not only the in-cylinder pressure pulse but also piston slaps, valve train motion, fuel injector pulses and other engine structural vibrations, which have little influence on knock characteristics, but they easily conceal slight knock. Even though an advanced knock module can be installed to reduce background noise, the knock module requires expertise to tune the frequency band, central frequency and gains. In addition, it is difficult for experts to determine the optimal parameters of the knock module to filter out background noise under difficult time-varying conditions.

The vibration signal detection method uses an accelerometer to detect knock characteristics by measuring the vibration acceleration of the cylinder block. Since this method has the advantages of easy installation, high reliability and low cost, it is commonly employed in real-time engine knock detection. Although using vibration signals to determine engine knock is more practical, a cylinder block vibration signal has a substantial amount of noise and signals from other vibration sources. Engine vibration signals cannot be applied to detect knock directly, and the original signals need to be processed using an accurate and effective signal denoising technique. Therefore, utilizing vibration signals for engine knock detection is still a challenging task.

Engine knock detection is a complicated problem that includes signal denoising, feature extraction, and feature classification. In signal denoising, although signals using variational mode decomposition (VMD) [35] are separated into a series of intrinsic mode functions (IMFs), IMFs depend on the values of the balancing parameter and the number of modes that are adjustable, and the results may be inaccurate when the parameters are not set in place. Therefore, there is an urgent need to obtain the optimal values of the VMD parameters. To reduce the computational burden in the later stage, some nonlinear dynamic parameters, such as energy ratio and correlation coefficient, should be used to extract the IMFs that represent prominent features. However, they are dependent on the record length, which is usually difficult or even impossible to acquire, especially in online condition monitoring and diagnosis. An appropriate indicator is also needed to determine sensitive subcomponents to select and reconstruct important IMFs. During feature extraction, each feature extraction method extracts different independent and complementary information from the signals. Therefore, an ensemble system using multiple feature learning is proposed to achieve high classification accuracy. Many experiments are usually needed to test the availability and performance of an optimal feature combination in specific applications. In feature classification, machine learning methods play an important role in the performance of the final classification results. Traditional neural networks and support vector machines have been applied to fault classification [17, 30]. However, they suffer from issues, including the computational burden of the large-scale fault classifier and the sensitivity of hyperparameters.

The main motivation of this research is to find the best solution in theory and application. In this paper, a novel intelligence engine knock detection system using multiple feature based sparse Bayesian extreme learning machine (SBELM), genetic algorithm-based VMD (GA-VMD) and sample entropy is proposed, and the salient contributions of this paper are organized as follows:

1) The traditional engine knock detection system usually relies on one kind of feature extracted from engine vibrations. Considering that the combination of different feature spaces from the observations would achieve better performance than any base classifier, an ensemble system using multiple feature learning is proposed to achieve high classification accuracy.
2) To overcome the dependency of the appropriate values of the balancing parameter and the number of modes, GA-VMD is used to filter unavoidable noise, in which the genetic algorithm (GA) is applied to obtain the optimal parameters to enhance the noise reduction ability. When the original time domain signals are decomposed into a series of IMFs, IMFs with sample entropy higher than the mean are selected as sensitive subcomponents for signal reconstruction.
3) This work is the first to attempt applying multiple features captured from engine vibration signals and SBELM together for engine knock detection. In addition to addressing the computational burden issue of the large-scale fault classifier, the ex-

tracted features are trained by SBELM to overcome the sensitivity of hyperparameters in conventional machine learning algorithms.

4) A universal intelligence solution for the detection task and the integration of GA-VMD with sample entropy, time domain statistical analysis (TDSA) , alpha stable distribution (ASD), and SBELM are also proposed to build an effective engine knock detection system.

This paper is organized as follows. The related work is briefly reviewed in Section 2. Section 3.1 introduces the outline of the engine knock detection system. The design procedure of the detection system and the signal filtering method are presented in Section 3.2, the feature extraction technique is presented in Section 3.3, and the classification procedure, which involves multiple techniques, is presented in Section 3.4. The performance evaluations of the proposed detection system are given in Section 4. Finally, a conclusion is summarized in Section 5.

2 Related Work

We briefly review previous approaches related to engine knock detection.

2.1 Signal Denoising

Engine knock detection can be viewed as an engine fault detection problem that relies on the features captured from a signal. The signal may contain noise or it can be affected by other component vibrations, so the knock-related information contained therein is not easy to observe. Therefore, many efforts have been made to develop signal processing techniques [32], such as fast Fourier transform [15], short-time Fourier transform [27], continuous wavelet transform [7, 39], discrete wavelet transform [6, 34] and nonlinear wavelet transform [16]. The fast Fourier transform method converts a time domain signal into a frequency domain signal quickly, but it is not suitable for non-stationary signals, such as the knock signal, which experiences rapid changes in both time and frequency. The short-time Fourier transform is an alternative transform method for time-frequency analysis, but it has not been extensively used due to its low time resolution with a fixed window under high frequencies. The resolution issue has been solved by wavelet transforms. However, the application of a wavelet transform has been bound by its inherent defect, which is the limitation of the selection of a mother wavelet, and it is a nonadaptive transformation. Empirical mode decomposition (EMD) [31] is self-adaptable and decomposes a signal directly into several IMFs, which are defined as amplitude-modulated-frequency-modulated signals whose number of local extrema and zero-crossings differ at most by one [13]. For the phenomenon that mode mixing occurs repeatedly in EMD, ensemble EMD (EEMD), which decreases the chance of undue mode mixing to a certain extent, was proposed [5]. The IMF in EEMD is characterized as the mean of an ensemble of trials whereby a finite-amplitude white noise signal is added to the decomposed data in each trial; this approach increases the computational burden since the data size of IMF is equal to that of the raw data. In recent years, VMD has been introduced for noise analyses of rotating machines and as a fault diagnosis method that has shown very promising results [8, 28, 35]. Although signals are separated into a series of IMFs, IMFs depend on the values of the balancing parameter and the number of modes that are adjustable, and the results may be inaccurate when they are not set in place [3]. Therefore, an optimization method utilizing GA is proposed in this work to solve the problem of parameter optimization.

To reduce the computational burden in the later stages, some nonlinear dynamic parameters, such as energy ratio [38] and correlation coefficient, should be taken to extract the IMFs that represent prominent features. However, the reliable estimation of both parameters depends on very long datasets, which are usually difficult or even impossible to acquire, especially during online condition monitoring and diagnosis. Entropy is defined as the loss of information in a time series or signal, such that approximate entropy [36] and sample entropy [24] are created to measure the repeatability or predictability within a time series. Due to its self-matching problem, approximate entropy is heavily dependent on the record length, and its value is uniformly lower than expected for short records and lacks relative coherence. Sample entropy is less dependent on the time series length and is utilized in this work to select and reconstruct important IMFs.

2.2 Feature Extraction

The selection of the feature extraction algorithm is known to play an important role in determining the performance of the classification system. An ensemble system using multiple feature learning is proposed to achieve high classification accuracy. This is made by combining the classifiers that are trained on different feature sets. The idea of combining different feature spaces from the observations made, that is, the combination of classifiers in different feature spaces, is the most effective way of combining classifiers and usually presents better results than any base classifier [9]. This occurs because each feature extraction method extracts different independent and complementary information from the signal. For this purpose, a diverse set of feature extraction methods using different approaches, such as TDSA, ASD and multi-fractal detrended fluctuation analysis (MFDFA), are selected.

2.3 Feature Classification

After feature exaction, machine learning methods play an important role in the performance of the final classification results. Traditional neural networks and support vector machines have been applied to fault classification [10, 14]. Much practical evidence shows that the long training time has greatly restricted the efficiency of these algorithms. In recent years, extreme learning machines (ELMs) have been utilized for multi-class classification based on a single hidden layer feed-forward network (SLFN) [12]. Recent studies show that the learning speed of ELM is faster than that of traditional learning algorithms [17, 26], so ELM can be suitable for large-scale problems. The dependent parameter of ELM is the number of hidden neuron nodes, but the initial hidden node parameters are random. Considering the susceptibility caused by the number of hidden neurons in conventional ELM, there might be a large number of hidden neurons selected in the trained model due to the minimization of the training error while ranking neurons, resulting in a high computational cost. Instead of explicitly adding/deleting hidden neurons in the conventional sparse ELMs, SBELM automatically tunes most of the output weights to zeros with an assumed prior distribution, thus gaining sparsity and achieving very high generalization. Hence, SBELM requires less calculation and is more suitable as a large-scale fault classifier.

2.4 Previous Schemes

Knock detection is usually a complicated problem that requires a combination of multiple techniques. Some previous schemes provided effective solutions by exploiting different technologies and ensured the reliability of knock detection. Sound vibration signal processing was proposed in [25]. In [25], a combination of methods, such as pass high-frequency filter, normalized envelope function and regression, was used to describe knock patterns, and then, the Euclidean distance was used to determine the existence of a detonation and achieved an accuracy of approximately 95%. However, the linear filter and distance-based classifier have limited noise reduction and feature classification abilities, respectively. A knock characteristic detection method based on wavelet denoising and EMD was proposed in [4]. The results indicated that the knock detection accuracy was 97%. An approach for detecting engine knocks of various intensities based on the vibration signal of an engine block using VMD and semi-supervised local Fisher discriminant analysis was proposed in [3], and the classification rate for strong knocks was over 95%. As mentioned above, there is much room for improvement in the denoising performance and accuracy.

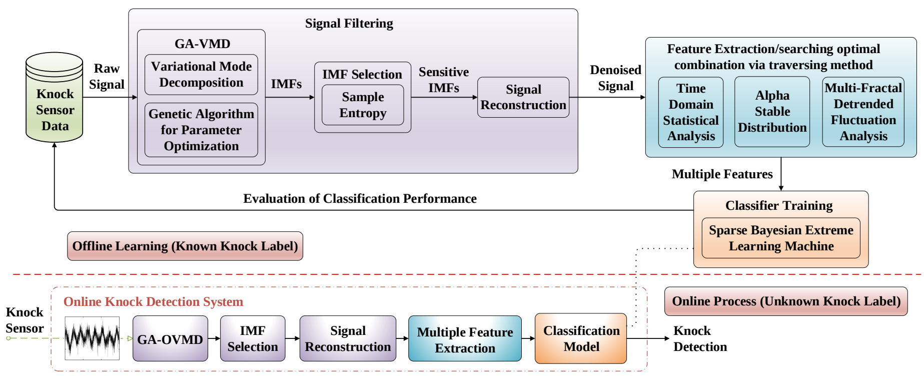


Fig. 1. Engine knock detection framework and project workflow

3 Designed of the Engine Knock Detection System

3.1 Outline of the Detection System

Motivated by the above general engine fault diagnostic requirements, a novel practical engine knock detection framework and project workflow are proposed in Fig.1. The proposed framework contains three main sections: signal filtering, feature extraction and classification. The GA-VMD method is developed to separate noise from the raw signal with a low computational burden compared with EEMD, where VMD is integrated with GA to achieve appropriate values of the balancing parameter and the number of modes. While the VMD converts the original signal into a series of IMFs, sensitive IMFs are then selected by sample entropy for further filtered signal reconstruction, and the unconsidered IMFs are removed. In terms of candidate feature extraction techniques before fault classification, TDSA, MFDFA, ASD and their possible combinations are tested to describe the distinguishable characteristics of the denoised signals. These features are trained by SBELM to establish a precision classifier. After the features of an unseen signal are fed to the trained classifier, a universal detection scheme is achieved to accurately identify engine knock online, such that the ECU can perform some actions, such as the retardation of the ignition in advance, to protect the engine.

3.2 Signal Filtering

GA-VMD For nonlinear and non-stationary time-frequency characteristics, GA-VMD is considered for signal filtering in the following work.

The goal of VMD is to decompose a real valued input signal f into a discrete number of sub-signals (*i.e.*, IMFs) u_k that have specific sparsity properties while reproducing the input. Here, the sparsity property of each mode is chosen to be its bandwidth in the spectral domain. In other words, we assume the kth mode to be mostly compact around a center pulsation ω_k, which is to be determined along with the decomposition.

To assess the bandwidth of a mode, the following scheme is proposed. (i) For each mode u_k, the associated analytic signal is computed by means of the Hilbert transform to obtain a unilateral frequency spectrum. (ii) For each mode, the frequency spectrum of the mode is shifted to the "baseband" by mixing an exponentially tuned value with the respective estimated center frequency. (iii) The bandwidth is now estimated through the Gaussian smoothness of the demodulated signal, *i.e.*, the squared L_2−norm of the gradient. The resulting constrained variational problem is given as follows,

$$
\begin{aligned}
&\min_{\{u_k\},\{\omega_k\}} \sum_{k=1}^{K} \left\| \partial_{\mathrm{t}} \left[\left(\delta(\mathrm{t}) + \frac{\mathrm{j}}{\pi \mathrm{t}} \right) * u_k(\mathrm{t}) \right] \exp(-\mathrm{j}\omega_k \mathrm{t}) \right\|_2^2 \\
&\text{s.t.} \sum_{k=1}^{K} u_k(\mathrm{t}) = f(\mathrm{t})
\end{aligned}
\tag{1}
$$

where t is the time script, δ is the Dirac distribution and $*$ denotes convolution. $\{u_k\} := \{u_1, \ldots, u_K\}$ and $\{\omega_k\} := \{\omega_1, \ldots, \omega_K\}$ are shorthand notations for the sets of all modes and their center frequencies, respectively. $k = 1, 2, \ldots, K$ and K is the number of modes of the intrinsic mode components.

The solution to Eq.(1) can be easily achieved via an unstrained optimization problem using the augmented Lagrangian method

$$\begin{aligned}\mathcal{L}(\{u_k\},\{\omega_k\},\lambda) := a\sum_{k=1}^{K}\left\|\partial_{\mathrm{t}}\left[\left(\delta(\mathrm{t})+\frac{\mathrm{j}}{\pi \mathrm{t}}\right)u_k(\mathrm{t})\right]\exp(-\mathrm{j}\omega_k t)\right\|_2^2 \\ +\left\|f(\mathrm{t})-\sum_{k=1}^{K}u_k(\mathrm{t})\right\|_2^2+\left\langle\lambda(\mathrm{t}), f(\mathrm{t})-\sum_{k=1}^{K}u_k(\mathrm{t})\right\rangle\end{aligned} \tag{2}$$

where a is the balancing parameter of the data-fidelity constraint, and λ is the Lagrange multiplier. An alternating direction method of multipliers is adopted to solve Eq.(2). The estimated modes u_k and the corresponding updated center frequency ω_k in the frequency domain can be achieved as follows:

$$u_k^{n+1}(\omega) = \frac{\tilde{f}(\omega) - \sum_{i<k} u_i^{n+1}(\omega) - \sum_{i>k} u_i^{n}(\omega) + \lambda^n(\omega)/2}{1 + 2a(\omega - \omega_k^n)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |u_k^{n+1}(\omega)|^2 d\omega} \tag{4}$$

where $\tilde{f}(\omega) := 1/\sqrt{2\pi}\int_{\mathbb{R}} f(\mathrm{t})\exp(-\mathrm{j}\omega \mathrm{t})d\mathrm{t}$ with $\mathrm{j}^2 = -1$, is the Fourier transform of the signal $f(\mathrm{t})$. The Lagrangian multiplier is updated as:

$$\lambda^{n+1}(\omega) = \lambda^n(\omega) + \tau_0\left(\tilde{f}(\omega) - \sum_k u_k^{n+1}(\omega)\right) \tag{5}$$

where τ_0 is the update parameter.

However, the values of the balancing parameter a and the number of modes K in Eq.(2) need to be predefined based on experience. For small values of a, one or more additional modes comprise noise. For large values of a, the essential parts of the signal are shared by at least two distinct modes, and their center frequencies overlap resulting in mode duplication. In addition, when the value of K is set too large, tampering features impede the accuracy of signal filtering, and essential intrinsic mode components are missed when the value of K is set too small. Additionally, the computational load can also be large due to the size of the data and a large mode number. Therefore, it is necessary to optimize those values to achieve satisfactory performance.

In the existing optimization techniques, many sequential search techniques are based on greedy methods. They are not suitable for global optimality but acceptable for local optimality. For instance, orderly searches consist of forward and backward selection. However, orderly forward and backward search techniques are not only more computationally expensive but also cannot perform undo processes, such as deleting or inserting features. In recent years, a novel memetic GA method for solving the traveling salesman problem was proposed in [1]. An application of GA and fuzzy goal programming to solve congestion management problems was proposed in [22]. The GA technique is based on evolutionary theory and the random search method. In this case, randomness is added to the search process to avoid local optima. GA is reliable and widely used in the optimization of artificial neural network parameters or signal processing algorithm parameters [28, 37]. Therefore, GA is introduced in this work to obtain the optimal values of the VMD parameters. For the optimization of signal processing parameters, the entropy concept is applied to the GA-VMD algorithm. In theory, a smaller entropy value leads to stronger properties and a clear signal distribution. The minimum envelope spectrum entropy value (MESEV) is proposed as the fitness function of the optimization and is obtained by the following steps:

(i) The Hilbert transform of an IMF signal, which is further described as a time series $\{u_k(t)\}$, can be expressed by

$$h_k(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u_k(t)}{t-\tau} d\tau \tag{6}$$

where $t = 1, 2, , \ldots, N$, and N is the length of the signal.

(ii) The envelope of the signal $u_k(t)$ is:

$$E_k(t) = \sqrt{u_k^2(t) + h_k^2(t)} \tag{7}$$

(iii) The envelope $E(t)$ is normalized as follows:

$$N_k(t) = \frac{E_k(t)}{\sum_{t=1}^{N} E_k(t)} \tag{8}$$

(iv) The envelope spectrum entropy value after normalization is:

$$V_k = -\sum_{t=1}^{N} N_k(t) \ln N_k(t) \tag{9}$$

(v) The MESEV is:

$$\langle a, K \rangle = \arg\min\{V_k\} \tag{10}$$

The proposed GA-VMD method is summarized in Fig. 2. The initial ranges for parameters a and K are assigned according to the actual situation at the beginning of the process. Then, GA-VMD initializes the population of GA and calculates the MESEV of each IMF. The operators in GA are compared to determine whether the current MESEV is the minimum. If not, the population is updated by new individuals until the minimum is reached. MESEV is used as a fitness function so that the iteration is stopped when the minimum MESEV converges to a stable constant or reaches the preset number of iterations. The values of a and K at the minimum MESEV are the optimal values.

Sample Entropy IMF selection methods that are commonly used in VMD are presented in this work to select and reconstruct important IMFs. Sample entropy is investigated to determine sensitive subcomponents.

Even though a higher energy ratio can reflect the fault-related information, faults usually appear at a low energy ratio. Noise always exists in raw signals and may cause incorrect IMF selections. By defining $N - m + 1$ templates, each of size m, which are

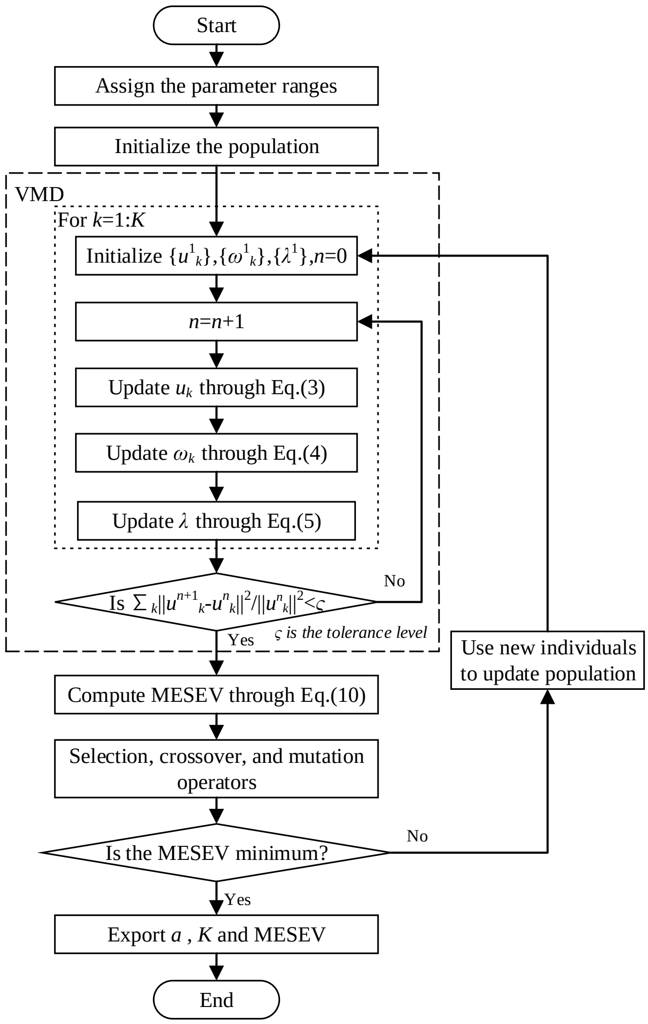


Fig. 2. Flowchart of the GA-VMD method

composed as $F^m(t) = [f(t), f(t+1), \ldots, f(t+m-1)]$, $U_k^m(t) = [u_k(t), u_k(t+1), \ldots, u_k(t+m-1)]$, $t = 1, \ldots, N-m+1$. The distance $d[F^m(t), U_k^m(t)]$, between $F^m(t)$ and $U_k^m(t)$, is computed as $d[F^m(t), U_k^m(t)] = \max\left[|f(t+j) - u_k(t+j)|\right]$, $j = 0, \ldots, m-1$. The sample entropy (SampEn) [24] is different from the energy-based method, which is expressed as:

$$\mathrm{SampEn}_k = \ln\left[\frac{\sum_{j=1}^{N-m+1} D_k^m(j)}{N-m+1} - \frac{\sum_{j=1}^{N-(m+1)+1} D_k^{m+1}(j)}{N-(m+1)+1}\right] \tag{11}$$

where $D_k^m(j) = \frac{N_k^m(j)}{N-m+1}$ is the probability that $U_k^m(t)$ matches $F^m(t)$, and $N_k^m(j)$ is defined as the number of template matches, *i.e.*, the number of $d[F^m(t), U_k^m(t)] < r$. In [23], Pincus suggested that the value of the threshold r should be selected between 0.1 and 0.25 and multiplied by the standard deviation of the raw signal and that m should be equal to 1 or 2. The IMFs with values higher than a preset threshold are chosen as the sensitive IMFs to reconstruct the denoised signal.

Remark 1. The above selection algorithms are used to determine the sensitive subcomponents from all IMFs, and the sensitive IMFs can reflect the knock features. The main pure signal is then reconstructed from the selected IMFs, *i.e.,* $\hat{f}(t) = \sum_{\mathcal{P}=1}^{\mathcal{K}} u_{\mathcal{P}}(t)$, where $u_{\mathcal{P}}(t)$ is the $\mathcal{P}$th sensitive IMF decomposed by VMD, and $\mathcal{K}$ is the number of sensitive IMFs.

3.3 Feature Extraction

In this section, a brief description of the three main feature sets used in the proposed multiple feature learning system is given.

Time Domain Statistical Analysis Traditionally, machinery signals were usually extracted by TDSA [29]. These statistical features describe the characteristics of a signal by a direct calculation with simple computations. Features such as standard deviation, root mean square, peak, skewness, kurtosis, crest factor, shape factor and impulse factor are employed in this work.

Alpha Stable Distribution ASD is suitable for describing random signals that have highly non-Gaussian distributions and heavy tails [33]. In ASD, the probability density function (PDF), which is utilized for describing the statistical characteristics of data, can be determined by the four parameters α, β, γ and δ. These parameters are usually expressed by their characteristic functions,

$$\phi(\mathrm{t}) = \exp\left(\mathrm{j}\delta\mathrm{t} - \gamma|\mathrm{t}|^{\alpha}\left[1 + \mathrm{j}\beta\,\mathrm{sign}(\mathrm{t})\theta(\mathrm{t},\alpha)\right]\right) \tag{12}$$

where $\theta(\mathrm{t},\alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ \left(\frac{2}{\pi}\right)\log|\mathrm{t}| & \alpha = 1 \end{cases}$. In this work, four parameters (α, β, γ and δ) are used to describe the different characteristics as features for further classification.

Multi-Fractal Detrended Fluctuation Analysis Detrended fluctuation analysis (DFA) is a fractal scaling method for perceiving long-range correlations in noisy and non-stationary time sequences. However, DFA is a mono-fractality method and is barely able to deal with multi-fractality nonlinear time series in dynamical mechanisms. Therefore, MFDFA was proposed for multi-fractality non-stationary time series analysis by extending the theory of DFA [18]. MFDFA has been verified in revealing the dynamic behavior hidden in multi-scale non-stationary signals and is described as follows.

The processed bounded time series $\{\hat{f}(1), \ldots, \hat{f}(t)\}$ is converted into an unbounded time series $\{\mathcal{F}(1), \ldots, \mathcal{F}(t)\}$ by a cumulative sum as follows:

$$\mathcal{F}(t) = \sum_{i=1}^{t}(\hat{f}(i) - \bar{f}(t)) \tag{13}$$

where $\bar{f}(t)$ is the mean of the time series $\{\hat{f}(1), \ldots, \hat{f}(t)\}$. Then, $\mathcal{F}(t)$ is divided into N_p non-overlapping segments with equivalent lengths p, where $N_p \equiv \text{int}(N/p)$. If N cannot be divided by p, the remaining part of the profile may be truncated. To retain with this unused part, the same process is implemented from the opposite end, and $2N_p$ segments are derived. For segment $l = 1, \ldots, N_p$, the least square of $F^2(p, l)$ is calculated as

$$F^2(p,l) = \frac{1}{p}\sum_{i=1}^{p}\Big(\mathcal{F}((l-1)p+i) - f_l(i)\Big)^2 \tag{14}$$

For segment $l = N_p + 1, \ldots, 2N_p$,

$$F^2(p,l) = \frac{1}{p}\sum_{i=1}^{p}\Big(\mathcal{F}(N-(l-N_p)p+i) - f_l(i)\Big)^2 \tag{15}$$

where $f_l(i)$ is a fitting polynomial in the lth segment. Different orders of the polynomial result in different eliminating trends from the profile. The qth order fluctuation function can be obtained by the average over all segments

$$F_q(p) = \left(\frac{1}{2N_p}\sum_{l=1}^{2N_p}\Big(F^2(l,p)\Big)^{q/2}\right)^{\frac{1}{q}} \tag{16}$$

where q is any real value except zero. Using different time scales of p, the scaling behavior of the fluctuation functions can be determined by analyzing the logarithmic relationship of $F_q(p)$ versus p for each q,

$$F_q(p) \propto p^{H(q)} \tag{17}$$

The relationship between the generalized Hurst exponent $H(q)$ and the scaling exponent $\tau(q)$ is as follows:

$$\tau(q) = qH(q) - 1 \tag{18}$$

The singularity exponent h_q and the multi-fractal singularity spectrum D_q are selected as the features and expressed as

$$h_q = \tau'(q) = H(q) + qH'(q) \tag{19}$$

$$D_q = qh_q - \tau(q) = q[h_q - H(q)] + 1 \tag{20}$$

where $H'(q)$ represents the derivative of $H(q)$ with respect to q. The Hölder exponent h_q characterizes the strength of the singularity, and D_q represents the Hausdorff dimension of the fractal subset with the exponent h_q, which are utilized to describe the different characteristics.

Remark 2. The three feature extractors describe the features from three aspects, and have multiple forms of arrangements and compositions. Time domain features have been proven to be effective for degradation monitoring and failure prognostics in the existing literatures. MFDFA is able to characterize the internal dynamics mechanism of fault signals and to detect slight changes in complex environments. The widely used ASD method has good robustness in the modeling of pulse shape in non-Gauss signals.

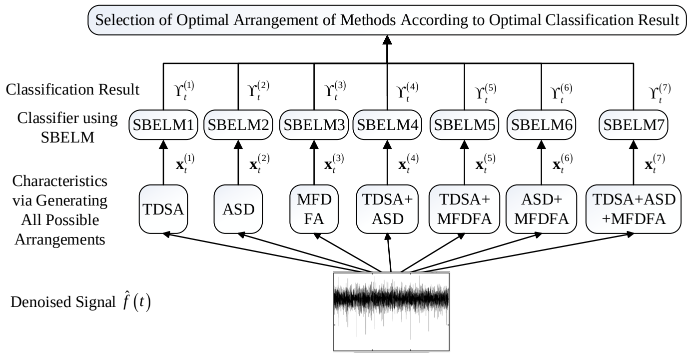


Fig. 3. Multiple feature learning process

Remark 3. The above feature extraction techniques, including TDSA, MFDFA, ASD, and their possible arrangements (*i.e.*, combinations), as shown in Fig.3, are tested to describe the distinguishable characteristics of the denoised signals. The optimal arrangement for finalizing the design of the feature extraction approach, as shown in Fig.4, is determined according to the optimal classification results obtained through the SBELM classifiers, which are described in the following section.

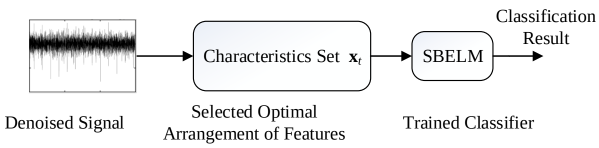


Fig. 4. Final knock detection system

3.4 Sparse Bayesian Extreme Learning Machine for Engine Knock Detection

The SBELM classifier is trained on data $(\mathbf{x}, \mathrm{T})$, which contain the above characteristics of any one arrangement and the known knock label. It is well known that neural network methods have been used successfully for fault diagnoses, and recently, a family of ELMs have been developed for training an SLFN with fast learning speeds and good generation performance. However, the execution time of ELM is quite unstable and depends on the number of hidden neurons (network size). Although a kernel-based ELM (KELM) that does not require hidden neurons and tends to provide better accuracy than basic ELM has been proposed, it suffers from large model size issues when the size of the training dataset is large. Before the development of ELM, relevance vector machine (RVM) was also available. RVM can train the kernel machine on a dataset and automatically prune the irrelevant basis elements to gain sparsity. To reduce the sensitivity of the number of hidden neurons in conventional ELM, SBELM was proposed, and it combines the advantages of the low computational load of ELM and the small weight and good prediction posterior probability of RVM. Reference [17] showed that when the number of hidden nodes is over 50, the classification accuracy could remain stable. This feature makes it more suitable as a large-scale fault classifier. The SBELM algorithm can be explained as follows.

The output weight of SBELM is learned by the Bayesian method instead of using the Moore-Penrose generalized inverse of the matrix [2]. The hidden layer output $\mathrm{H} = [h_1, \ldots, h_t, \ldots, h_N]^T$ becomes the input of SBELM, where $h_t \in R^L$ is the hidden feature mapping with respect to input $\mathbf{x}_t \in R^L$, L is the number of characteristics of the optimal arrangement, and N is the number of classifier outputs. Each training sample $\mathbf{x}_t$ from the extracted features can be treated as an independent Bernoulli case.

Using iterative reweighted least squares to find the Laplace mode $\hat{W}$ is efficient; hence the gradient ∇E and Hessian matrix ϕ must be computed:

$$\nabla E = \nabla_W \ln\{P(\mathrm{T}|\mathrm{W}, \mathrm{H})P(\mathrm{W}|\boldsymbol{\alpha})\} = \mathrm{H}^T(\mathrm{T} - \mathrm{Y}) - \mathrm{AW} \tag{21}$$

$$\phi = \nabla_W \nabla_W \ln\{P(\mathrm{T}|\mathrm{W}, \mathrm{H})P(\mathrm{W}|\boldsymbol{\alpha})\} = -(\mathrm{H}^T\mathrm{BH} + \mathrm{A}) \tag{22}$$

where $\mathrm{W} = (w_1, \ldots, w_m, \ldots, w_L)^T$ is the hidden layer matrix. $\mathrm{T} = (\mathcal{T}_1, \ldots, \mathcal{T}_t, \ldots, \mathcal{T}_N)^T$, $\mathcal{T}_i \in \{0, 1\}$ is a target output vector. $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_L]^T$ is the independent prior in relation to each w_m, and some values of w_m are regulated to zero by adaptive rectangular decomposition (ARD) to select important hidden neurons. $\mathrm{Y} = (y_1, \ldots, y_N)^T$, where $y_t = \sigma(h_t, w_t)$, $\mathrm{A} = \mathrm{diag}(\boldsymbol{\alpha})$ and B is a diagonal matrix, where $b_t = y_t(1 - y_t)$. Subsequently, $\hat{\mathrm{W}}$ can be obtained by

$$W_{new} = W_{old} - \phi^{-1}\nabla E = (\mathrm{H}^T\mathrm{BH} + \mathrm{A})^{-1}\mathrm{H}^T\mathrm{B}\hat{\mathrm{T}} \tag{23}$$

where $\hat{\mathrm{T}} = \mathrm{HW} + \mathrm{B}^{-1}(\mathrm{T} - \mathrm{Y})$. The center $\hat{\mathrm{W}}$ and covariance matrix $\varSigma$ of the Gaussian distribution are

$$\varSigma = (\mathrm{H}^T\mathrm{BH} + A)^{-1} \quad \text{and} \quad \hat{\mathrm{W}} = \varSigma\mathrm{H}^T\mathrm{B}\hat{\mathrm{T}} \tag{24}$$

As a result, $\ln\{P(\mathrm{T}|\mathrm{W}, \mathrm{H})P(\mathrm{W}|\boldsymbol{\alpha})\} \propto N(\hat{\mathrm{W}}, \varSigma)$ is formed and the log marginal likelihood $L(\boldsymbol{\alpha}) = \ln P(\mathrm{T}|\boldsymbol{\alpha}, H)$ can be computed by setting $L(\boldsymbol{\alpha})$ to zero as follows:

$$\frac{\partial L(\boldsymbol{\alpha})}{\partial \alpha_m} = \frac{1}{2\alpha_m} - \frac{1}{2}\varSigma_{mm} - \frac{1}{2}\hat{w}_m^2 = 0 \rightarrow \alpha_m^{new} = \frac{1 - \alpha_m\varSigma_{mm}}{\hat{w}_m^2} \tag{25}$$

By setting the initial values of w_m and α_m, $\hat{W}$ and $\varSigma$ are updated by Eq.(24) and the values of α_m are updated by substituting α_m and $\varSigma$ into Eq.(25). The marginal likelihood function is iterated to the maximum value until the convergence criterion is met.

In summary, the whole learning procedure of the fault diagnosis scheme is given below. Given the knock label $\mathcal{T}_t$ and the training denoised signal $\hat{f}(t)$, the training procedure is shown as follows.

Training procedure

(i) Extract the characteristic data $\mathbf{x}_t^{(r)}$ via generating all possible arrangements of three feature extraction methods from the denoised training signal $\hat{f}(t)$, $r = 1, \ldots, 7$

(ii) For each arrangement,
Initialization: randomly generate input weights and calculate the output of hidden layer H, W = $\mathbf{0}$, $\boldsymbol{\alpha} = 10^{-5}\mathbf{1}$
Step 1: Estimation of output weights W

(a) Set the initial value $\Sigma = \mathbf{0}$, and define an intermediate variable $g = \mathbf{0}$

(b) Sequentially calculate the mapping of every input $\mathbf{x}_t^{(r)}$ to h_t with random ELM hidden weights
For $t = 1 : N$
$\epsilon = \epsilon + y_t(1 - y_t)h_t^T h_t$
$g = g + (-1)(\mathcal{T}_t - y_t)h_t^T$
End for

(c) $\Sigma = (\epsilon + \text{diag}(\boldsymbol{\alpha}))^{-1}$, $\nabla E = g + \text{diag}(\boldsymbol{\alpha})W$

(d) Find step size λ with line search method, W = $W - \lambda\Sigma^{-1}\nabla E$

(e) If $norm(\nabla E)$ is under a predefined gradient tolerance, then go to Step 2. Otherwise, go to Step 1.
Step 2: Estimation of hyperparameter $\boldsymbol{\alpha}$.

(f) For every α_m
$\alpha_m = (1 - \alpha_m \Sigma_{mm}^{-1})/w_k^2$
End for
Step 3: Pruning nodes

(g) If α_m >predefined maximum
prune α_m, w_m, $H(:, m)$, $L = L - 1$
End if

(h) If the absolute difference between two successive logarithm values of α_m is lower than given tolerance, then stop. Otherwise, repeat **Step 1 to Step 3.**

(iii) Calculate the classifier results of each arrangement, and select the optimal arrangement.

Testing procedure For each denoised signal $\hat{f}(t)$,

(i) Extract the characteristic data $\mathbf{x}_t$ via selected optimal arrangement from the denoised signal $\hat{f}(t)$.

(ii) Calculate the output of the related classifier, whose parameters are inherited from training procedure.

4 Experiment and Evaluation

4.1 Experimental Setup

To test and train the proposed framework, a test rig is designed to collect the raw engine data and is presented below.

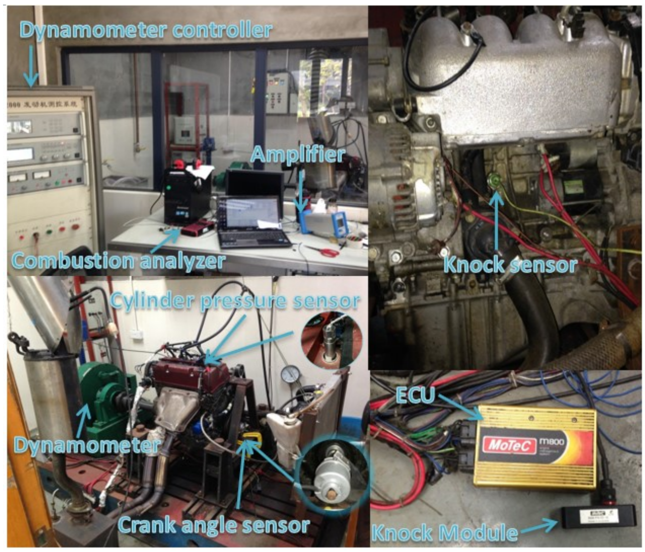


Fig. 5. Test rig

A *Honda K20A Type-R* engine, which is a four-stroke, four-cylinder spark-ignition engine, is utilized as the test rig as shown in Fig. 5. The research octane number of the fuel is 98, which was purchased from a regular gas station. The experimental setup as shown in Fig. 6 can be divided into three main sections. The first section contains the ECU, the engine and relative peripheral sensors, where the raw data are collected via a knock sensor. The second section contains the dynamometer and its control system for varying the loading condition of the engine. The third section contains the combustion analyzer with an in-cylinder pressure sensor, which is used to detect whether knock exists in the experiment. The data collected by the in-cylinder pressure can validate the result of the proposed system. The main components are as follows:

Electronic Control Unit A *MoTeC M800* programmable ECU controls the engine by monitoring sensor signals and adjusting the outputs based on the look-up tables. The ECU can control the spark timing, fuel injection time and engine temperature. In this work, the injection time and ignition timing are important for ECU control. During the experiment, the injection time and ignition timing at different engine speeds and loads can be adjusted through the fuel map and ignition map in the ECU, respectively. The fuel map mainly controls the air-fuel ratio or air ratio. To measure the air-fuel ratio/air ratio, a lambda sensor/oxygen sensor is installed in the exhaust pipe and used for measurement.

Dynamometer and Control System A *DW160* eddy-current dynamometer is used to apply the engine load and control the engine throttle for simulating different driving conditions. The dynamometer is coupled to the test engine.

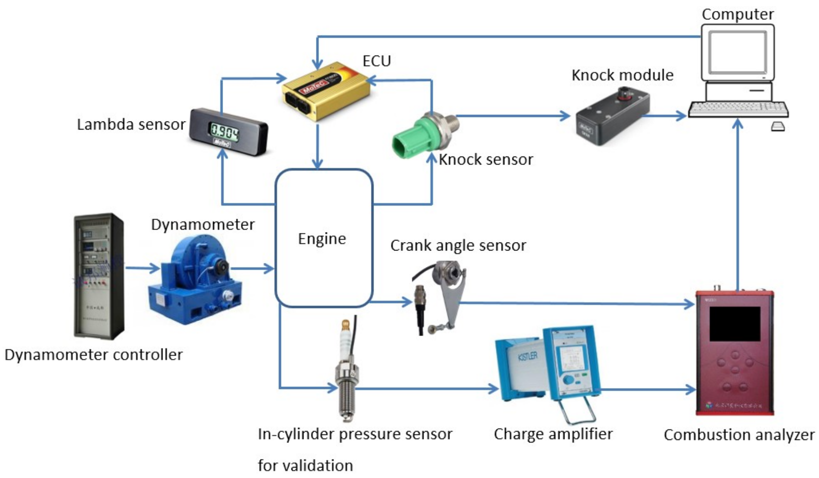


Fig. 6. Test rig setup

Combustion Analyzer An *MA3001* combustion analyzer, which was produced by PowerMAC Co., Ltd., is used to analyze the in-cylinder pressure and corresponding crank angle. The analyzer consists of two parts: (i) The crank angle sensor, which is mounted on the engine crankshaft terminal to measure the engine crank angle in the engine cycle. The sensor is used to convert the rotational speed and phase position of the crankshaft into a digital angle signal, which helps monitor the pressure wave for knock detection. (ii) A piezoelectric in-cylinder pressure sensor is employed to measure the in-cylinder combustion pressure for validation. The signal from the cylinder pressure sensor is then amplified by a charge amplifier. The crank angle signal and the amplified in-cylinder pressure signal are sent to the analyzer for pressure wave analysis. Before starting the experiment, the devices had to be calibrated. The calibrated range and sensitivity charge of the amplifier are set to $150bar$ and $-10.22pC/bar$ to match the in-cylinder pressure sensor. The mode of the amplifier is set to $0-10V$ according to the specification of the combustion analyzer. The voltage-pressure conversion coefficient of the combustion analyzer is set to 15, depending on the amplifier and the test engine torque. It is worth noting that the top dead center position needs to be calibrated when the crank angle sensor is installed on the test engine.

Data Collection and Analysis A software called *GoldWave* is installed on a computer to record the engine signals from the knock sensor. The signal is then passed to MATLAB to conduct signal filtering, feature extraction and classification.

4.2 Operating Conditions for Experiment Data Collection

To verify the proposed scheme, real engine data are recorded and analyzed. Since the fuel used in the experiment has a high-octane number, engine knock does not easily occur. To produce knock conditions under different driving conditions without damaging the engine in the laboratory, the engine is operated under two working conditions: i) low speed with high load conditions and ii) high speed with low load conditions. The engine load is provided by the dynamometer by applying opposite torque to the engine. The ignition timing is advanced gradually. The initial engine temperature before knocking is held at 85°C ± 5°C. The engine load, speed and air-fuel ratio are changed within a certain range. The combustion analyzer records the pressure wave pattern to determine the presence of engine knock so that the training and test data can be obtained. A total of 1800 sets of data are recorded according to different driving conditions, as shown in Table 1.

Table 1. Experimental data

Operation condition				Number of samples	Objective
Speed (rmp)	Load (Nm)	Air-fuel	Ignition Timing (°BTDC)		
1000±300	60 ± 5	1±0.5	10°±2.5 to 45°±2.5	320	Simulate a low speed and high load driving condition
2000±300	12 ± 5	1±0.5	10°±2.5 to 45°±2.5	990	Simulate a high speed and low load driving condition
3000±300	12 ± 5	1±0.5	10°±2.5 to 45°±2.5	490	Simulate a high speed and low load driving condition

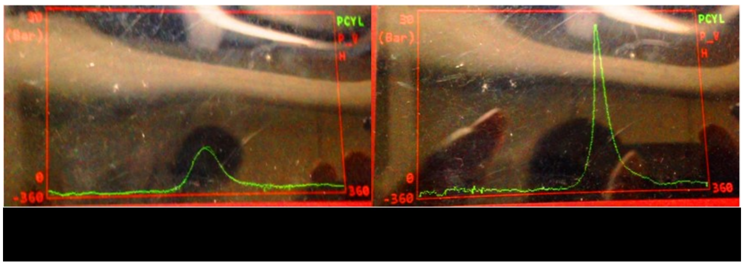


Fig. 7. Pressure versus crank angle under (left) non-knock and (right) knock conditions

At the beginning of the experiment, knock does not occur easily at idle speeds due to the high anti-knock quality of the fuel, even when the ignition timing is substantially advanced and the air-fuel ratio is enriched. Under this condition, the cylinder pressure wave pattern in the combustion analyzer is still smooth, as shown on the left-hand side of Fig.7. When the ignition advances and the engine load continues to increase, the shape of the pressure wave sharply increases. When the ignition timing and engine load are increased to a certain range, an obviously high and sharp pressure wave appears, indicating the existence of knock, as shown on the right-hand side of Fig.7. Therefore, it is not easy to generate a knock at a low engine speed with a high-octane fuel unless the engine load is high. Certainly, engine operating at a high engine speed under a high-octane fuel can generate knock easily under a low engine load. It is noteworthy that the combustion analyzer and in-cylinder pressure signal are not suitable for in-use vehicles

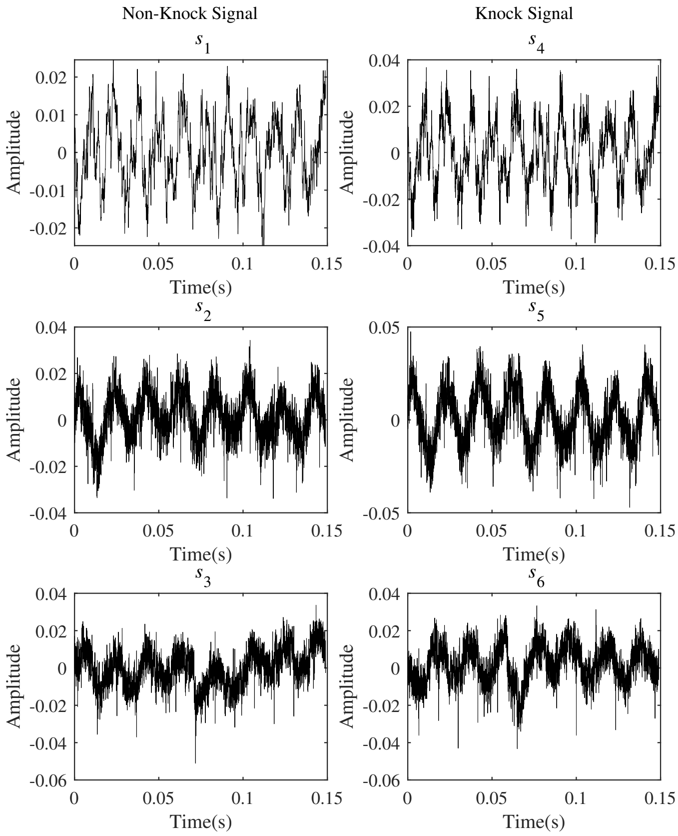


Fig. 8. Time domain engine vibration signals

due to their high costs, so they are used only for validation and labeling only. The actual knock detection signal is the engine vibration signal captured by the knock sensor.

The vibration signal collected by the knock sensor converts the shock of cylinder pressure into an electronic signal. For each driving condition, the raw signals are recorded for 0.15 seconds with a sampling rate of 48000 Hz. Therefore, each sample contains a time series with 7200 sampling points. Six randomly selected vibration signals from the 1800 sets of data shown in Table 2 are illustrated in Fig.8, where half of the signals are non-knock labeled signals and half are knock labeled signals. They are used as training dataset to train the classifiers. It can be observed from Fig.8 that the non-knock signals (s_1, s_2, s_3) are very difficult to manually distinguish from the knock signals (s_4, s_5, s_6). Therefore, the proposed framework is applied to remove noise from

Table 2. Experimental setup of the sample vibration signals

Vibration Signal		Speed (rpm)	Load (Nm)	Air-fuel ratio	Ignition timing (°BTDC)
Non-knock	s_1	1000	58.1	0.9	20
	s_2	2000	7	0.9	20
	s_3	3000	9	0.9	20
Knock	s_4	1000	58.1	0.9	40
	s_5	2000	12	0.9	40
	s_6	3000	15	0.7	42

vibration signals and detect knock. The experimental data and program code in MATLAB are available at https://github.com/wangdai11/EKDS.

4.3 Results and Evaluation

Signal Filtering Signal filtering is the first step of the proposed framework, and it reduces noise from the raw vibration signals. VMD converts the raw signals into a series of IMFs. Sample entropy is employed in the proposed signal processing methods to remove the insensitive IMFs. For comparison, signal s_6 is used as an example in this section to evaluate the filtering ability of the proposed GA-VMD.

IMFs of VMD depend on the adjustable parameters a and K, which are inaccurate when the parameters are set inappropriately. Therefore, GA is proposed to obtain the appropriate values for a and K. The parameters of GA are set as follows: population size=50, number of generations=200, mutation rate=0.01, mutation percentage of the population=0.2, and crossover percentage of the population=0.8. The input ranges of a and K are set to $[100, 10000]$ and $[2, 20]$ respectively. After 50 runs of GA, the average values are $a = 1463$ and $K = 9.9$ respectively. Therefore, a and K are set to 1500 and 10.

Fig. 9 illustrates an example that shows the influence of different values of a and K on signal filtering. When a is set too large or when K is set inappropriately, some knock resonant frequencies (Fig. 9f, 9h, 9j, and 9l) cannot be clearly displayed compared with Fig. 9b. Choosing sample entropy as the IMF selection method due to the best noise reduction ability, Fig. 9c and Fig. 9d show the GA-VMD results. Fig. 9c, Fig. 9d and Appendix A show that only GA-VMD can clearly reflect all the resonant frequencies.

The results of using VMD and different IMF selection methods for signal s_6 are shown in Fig. 10 and Table 3. Each method takes the threshold T to select the appropriate IMFs for signal reconstruction, where $T = \frac{\sum_{n=1}^{K} IMF_K}{K}$ and K is the total number of IMFs. The IMFs with values higher than the threshold are chosen and highlighted in red in Table 3. The selected IMFs are reconstructed into a denoising signal and the envelope spectrum of the filtered signals is used to identify the knock resonant frequency. Fig. 11 shows the envelope spectrum of the GA-VMD noise reduction under different IMF selection methods. Fig. 11h and Appendix B show that only the sample entropy can reflect the knock resonant frequencies as shown in Fig. 11e. This further indicates that the sample entropy approach has good noise reduction and signal reconstruction abilities.

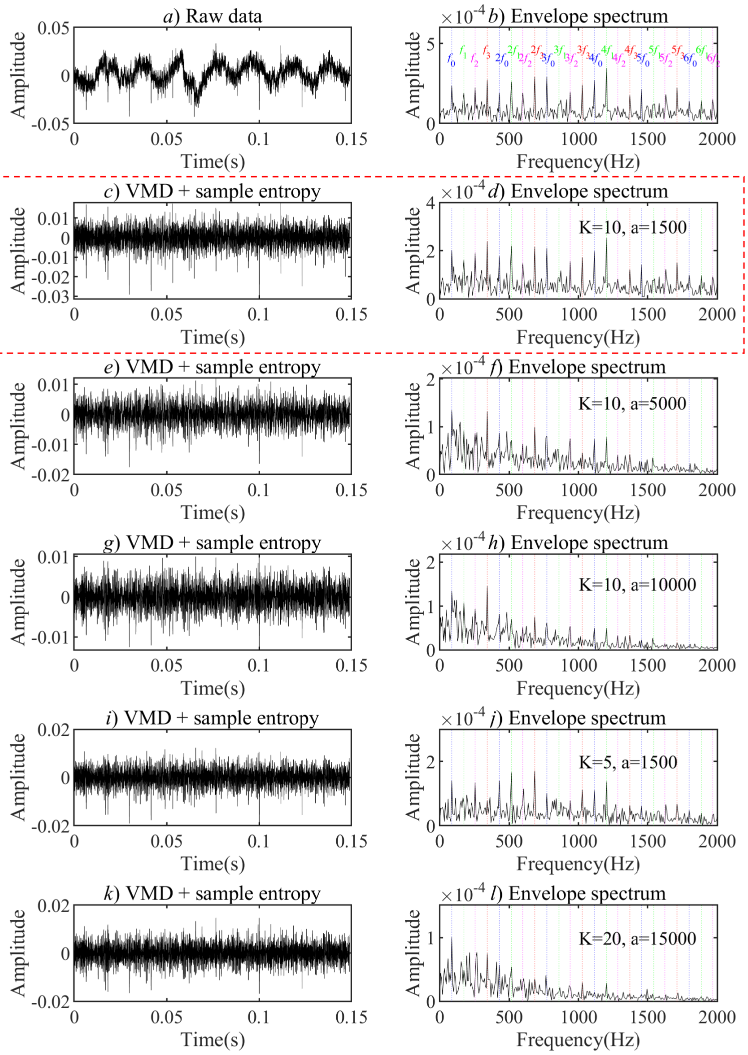


Fig. 9. Noise reduction ability under different values of a and K

Feature Extraction Feature extraction, a pretreatment for machine learning methods, is the second step of the proposed knock detection method. The applications of TDSA, ASD and MFDFA are used for extracting cognizable features from the filtered signals.

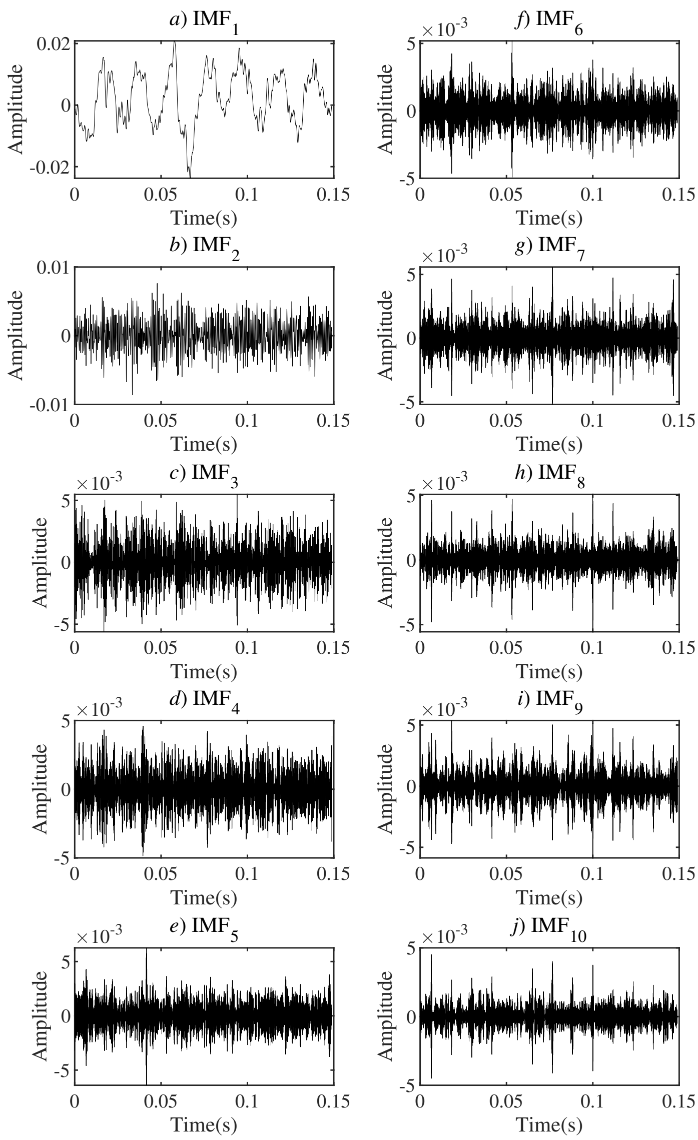


Fig. 10. IMFs obtained based on GA-VMD

Table 3. Results of GA-VMD with different IMF selection methods

s_6	Correlation Coefficient	Energy Ratio	Sample Entropy
IMF_1	0.7790	0.5632	0.0645
IMF_2	0.3963	0.0744	0.5218
IMF_3	0.3051	0.0356	0.6086
IMF_4	0.2535	0.0236	0.5841
IMF_5	0.2338	0.0215	0.5775
IMF_6	0.2086	0.0162	0.5748
IMF_7	0.1954	0.0147	0.5362
IMF_8	0.1875	0.0128	0.5934
IMF_9	0.1934	0.019	0.5911
IMF_{10}	0.1442	0.0084	0.5979
T	0.2897	0.0790	0.5260

Each extracted feature can compress a large number of time series data into specific numbers. These specific numbers representing meaningful features are then used to establish a classification model for knock detection.

Table 4 shows the TDSA features of 24 randomly selected engine vibration signals under different conditions, including mean y_{mean}, standard deviation y_{std}, root mean square y_{rms}, peak y_{peak}, skewness y_{skew}, kurtosis y_{kurt}, crest factor y_{crf} and y_{clf}, shape factor y_{sf} and impulse factor y_{if}, which are created under different ignition timing and loading conditions. In Table 4, the sample signals A_1 to A_8 are at 1000 rpm, B_1 to B_8 are at 2000 rpm and C_1 to C_8 are at 3000 rpm. These statistical features can be used to separate knock data from non-knock data. Therefore, these statistical features are kept for the inputs of the classifiers.

The ASD algorithm is a feature extraction method that emphasizes the characteristic parameters α, β, γ, and δ. The values of these parameters are self-generated by the wave patterns of the signal. The ASD characteristic parameters and the magnitudes of the PDF are different under knock and non-knock conditions as shown in Fig. 12. Therefore, the parameters α, β, γ, δ and h are selected as the inputs of the classifiers. Table 6 shows the five ASD parameters of the same 24 vibration samples (A_1 to A_8, B_1 to B_8 and C_1 to C_8) in Table 4.

Fig. 13 depicts that the knock data mainly lay between the large values of γ and α, but the non-knock data are dispersive. Most of the non-knock data have higher values of h and α than the knock data. In this case, most knock data can be separated from the non-knock data with this method.

MFDFA is another feature extraction approach that emphasizes the 3 points in the multi-fractal spectrum: i) the first points of the multi-fractal curves (h_{q_a},D_{q_a}); and ii) the end points of the multi-fractal curves (h_{q_b},D_{q_b}); and iii) the peaks of the multi-fractal curves (h_{q_0},1). The signal under various working conditions provides different spectra, as shown in Fig. 14. Table 7 shows the five multi-fractal parameters (h_{q_a}, D_{q_a}, h_{q_b}, D_{q_b}, and h_{q_0}) of the same 24 vibration samples (A_1 to A_8, B_1 to B_8 and C_1 to C_8). The distribution results of the multi-fractal parameters in Fig. 15 show that most of the knock data in Table 4 can also be separated from the non-knock data under GA-VMD. Therefore, MFDFA is also considered in this work.

The above three feature extraction methods are feasible and produce different separable features, so they are used in different combinations. In total, 20 features based

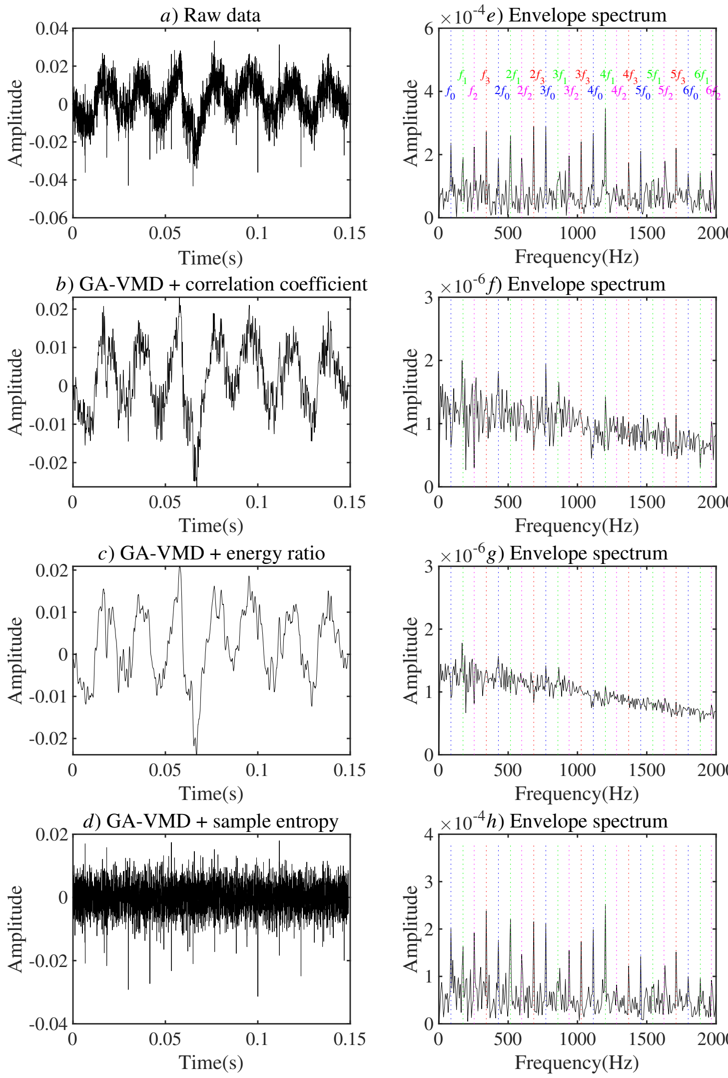


Fig. 11. Envelope spectrum of GA-VMD under different IMF selection methods

on TDSA, ASD and MFDFA methods are obtained, as shown in Table 5. The different characteristics of the knock data and non-knock data are then entered into the machine learning methods for building classifiers for diagnosis.

Table 4. Example of the TDSA result of GA-VMD+Sample entropy

		y_{mean}	y_{std}	y_{rms}	y_{peak}	y_{skew}	y_{kurt}	y_{crf}	y_{clf}	y_{sf}	y_{if}
Non-Knock	A_1	2.29×10^{-7}	9.87×10^{-3}	0.001	0.010	-0.112	4.508	5.318	8.438	1.315	6.991
	A_2	-4.19×10^{-8}	9.70×10^{-3}	0.001	0.010	0.124	5.063	5.842	9.351	1.322	7.721
	A_3	-4.80×10^{-7}	2.60×10^{-3}	0.003	0.011	0.017	3.000	4.002	5.913	1.253	5.014
	A_4	-3.40×10^{-7}	2.36×10^{-3}	0.002	0.010	0.004	3.046	4.108	6.124	1.258	5.169
	B_1	2.56×10^{-6}	2.42×10^{-3}	0.002	0.020	0.151	5.609	7.186	11.518	1.326	9.527
	B_2	8.60×10^{-7}	4.11×10^{-3}	0.004	0.030	-0.098	4.153	6.326	9.556	1.274	8.060
	B_3	2.94×10^{-6}	2.47×10^{-3}	0.002	0.022	-0.073	5.251	7.651	12.234	1.323	10.124
	B_4	2.01×10^{-6}	3.72×10^{-3}	0.004	0.036	-0.242	4.882	7.204	11.008	1.285	9.257
	C_1	5.28×10^{-7}	2.09×10^{-3}	0.002	0.029	0.103	7.403	10.688	17.640	1.354	14.470
	C_2	9.29×10^{-7}	3.87×10^{-3}	0.004	0.037	-0.273	4.680	7.435	11.426	1.289	9.582
	C_3	-6.60×10^{-6}	4.39×10^{-3}	0.004	0.037	-0.266	4.537	7.492	11.530	1.289	9.656
	C_4	-3.10×10^{-6}	4.39×10^{-3}	0.004	0.043	-0.308	4.789	7.037	10.916	1.296	9.122
Knock	A_5	-5.66×10^{-7}	4.11×10^{-3}	0.004	0.014	-0.095	2.994	5.204	1.253	1.253	4.408
	A_6	6.31×10^{-7}	3.28×10^{-3}	0.003	0.014	0.037	3.099	6.204	1.253	1.253	5.433
	A_7	-2.13×10^{-6}	3.29×10^{-3}	0.003	0.014	-0.068	3.159	6.535	1.263	1.263	5.509
	A_8	-5.12×10^{-7}	4.05×10^{-3}	0.004	0.017	0.054	3.642	6.534	1.278	1.278	5.506
	B_5	8.55×10^{-6}	5.29×10^{-3}	0.005	0.027	-0.053	3.604	7.890	1.279	1.279	6.662
	B_6	4.94×10^{-7}	4.91×10^{-3}	0.005	0.027	-0.041	3.789	8.380	1.281	1.281	7.027
	B_7	-1.52×10^{-6}	5.45×10^{-3}	0.005	0.035	-0.103	3.817	9.849	1.275	1.275	8.298
	B_8	8.95×10^{-6}	6.06×10^{-3}	0.006	0.030	-0.100	3.577	7.410	1.271	1.271	6.246
	C_5	9.29×10^{-7}	3.87×10^{-3}	0.004	0.029	-0.183	4.496	11.467	1.281	1.281	9.662
	C_6	-6.82×10^{-6}	5.21×10^{-3}	0.005	0.033	-0.192	4.196	9.614	1.283	1.283	8.068
	C_7	5.41×10^{-6}	6.16×10^{-3}	0.006	0.030	-0.070	3.528	7.338	1.278	1.278	6.156
	C_8	1.49×10^{-5}	5.33×10^{-3}	0.005	0.029	0.023	3.676	8.363	1.279	1.279	7.013

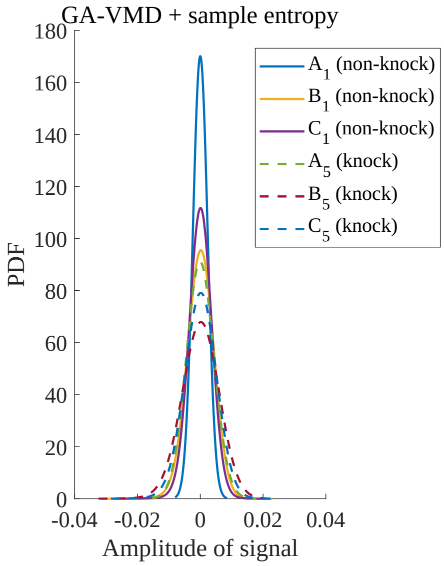


Fig. 12. PDF spectrum of different signals

Classification Classification is the last step of the proposed framework. The extracted features are learned using two other machine learning algorithms, where ELM and

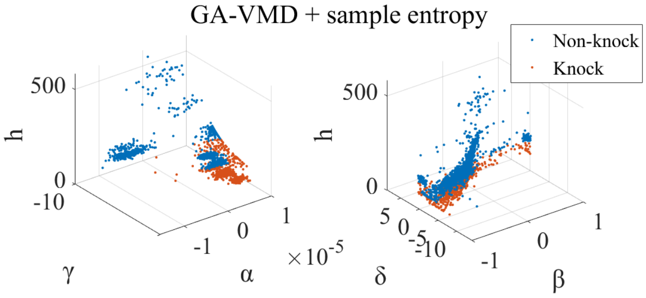


Fig. 13. ASD parameters

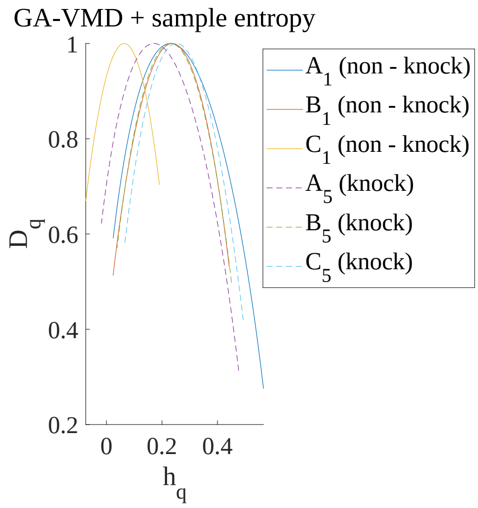


Fig. 14. Multi-fractal spectra of different signals

Table 5. Extracted features

Methods	Features	Total
TDSA	Mean, standard deviation, root mean square, peak, skewness, kurtosis, crest factor, clearance factor, shape factor, impulse factor	10
ASD	$\alpha, \beta, \gamma, \gamma, h$	5
MFDFA	$h_{qa}, D_{qa}, h_{qb}, D_{qb}, h_{q0}$	5

KELM are applied for comparison. To verify the statistical performance of the test results, we use bootstrapping for the dataset. Bootstrapping is a test or metric that relies on random sampling with replacement. The dataset is separated into two groups, non-knock data and knock data, wherein 900 sets are randomly selected as training data and the remaining 900 sets are used as test data. The division of the training and test

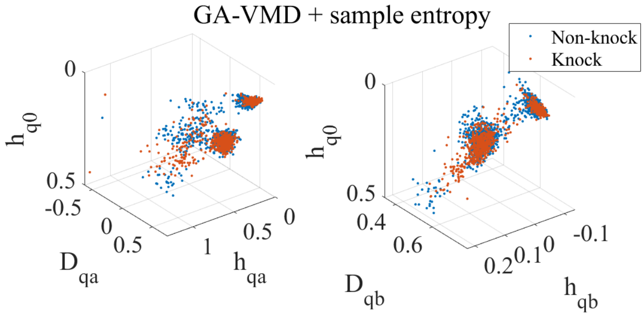


Fig. 15. MFDFA parameters

Table 6. ASD results with GA-VMD+Sample entropy

		α	β	γ	δ
Non-Knock	A_1	2.000	-1.000	1.90×10^{-3}	-2.86×10^{-5}
	A_2	1.957	-0.112	2.32×10^{-3}	1.60×10^{-5}
	A_3	1.847	0.154	1.02×10^{-3}	4.39×10^{-5}
	A_4	2.000	1.000	2.06×10^{-3}	6.37×10^{-6}
	B_1	1.968	-1.000	2.91×10^{-3}	-1.89×10^{-5}
	B_2	1.951	-0.448	2.95×10^{-3}	5.02×10^{-7}
	B_3	1.805	0.076	1.29×10^{-3}	1.91×10^{-6}
	B_4	1.977	0.609	2.63×10^{-3}	3.54×10^{-5}
	C_1	1.966	-1.000	2.96×10^{-3}	-3.73×10^{-5}
	C_2	1.793	-0.389	1.49×10^{-3}	-4.61×10^{-5}
	C_3	1.828	0.079	1.30×10^{-3}	1.28×10^{-5}
	C_4	1.795	0.529	1.55×10^{-3}	7.55×10^{-5}
Knock	A_5	1.990	-1.000	3.30×10^{-3}	-1.48×10^{-5}
	A_6	1.972	-0.173	3.12×10^{-3}	1.50×10^{-5}
	A_7	1.976	0.068	2.99×10^{-3}	3.95×10^{-5}
	A_8	1.988	1.000	3.33×10^{-3}	7.06×10^{-6}
	B_5	1.964	-1.000	4.05×10^{-3}	4.44×10^{-5}
	B_6	1.927	-0.523	4.08×10^{-3}	2.03×10^{-6}
	B_7	1.943	-0.305	4.52×10^{-3}	7.81×10^{-6}
	B_8	1.976	0.627	4.60×10^{-3}	-9.65×10^{-6}
	C_5	1.952	-1.000	3.56×10^{-3}	-5.84×10^{-5}
	C_6	1.926	-0.295	3.04×10^{-3}	-1.30×10^{-5}
	C_7	1.929	0.013	2.99×10^{-3}	1.93×10^{-5}
	C_8	1.962	0.163	4.08×10^{-3}	2.05×10^{-5}

datasets is presented in Table 8. The mean results are achieved after 10 repetitions and are shown in Table 9.

Table 8 shows that this knock detection problem is a binary classification problem. To select an appropriate classification method, the accuracies of the three machine learning methods are compared. For ELM and SBELM, the number of initial hidden neurons must be defined. The initial hidden neurons for ELM and SBELM are set to 200. For KELM, the kernel is introduced to the model; thus, the regularization parameter and kernel parameter have to be set. The kernel function of KELM is a radial basis function. The regularized parameter and the kernel parameter of KELM are set to 1.0. The test accuracies are shown in Table 9, and the best accuracy is highlighted in red. Ta-

Table 7. MFDFA results with GA-VMD+Sample entropy

		h_{q_0}	h_{qa}	D_{qa}	h_{q_b}	D_{q_b}
Non-Knock	A_1	0.098	-0.090	0.599	0.370	0.356
	A_2	0.215	0.012	0.602	0.584	0.162
	A_3	0.317	0.078	0.548	0.692	0.204
	A_4	0.230	0.019	0.590	1.099	-0.626
	B_1	0.050	-0.042	0.793	0.164	0.727
	B_2	0.058	-0.064	0.713	0.238	0.525
	B_3	0.202	0.010	0.593	0.439	0.449
	B_4	0.278	0.048	0.454	0.547	0.347
	C_1	0.048	-0.077	0.702	0.160	0.746
	C_2	0.062	-0.070	0.699	0.235	0.566
	C_3	0.244	0.061	0.579	0.443	0.547
	C_4	0.263	0.035	0.506	0.563	0.271
Knock	A_5	0.119	-0.070	0.600	0.455	0.231
	A_6	0.243	0.035	0.599	0.561	0.308
	A_7	0.340	0.105	0.520	0.693	0.243
	A_8	0.429	0.181	0.517	1.316	-0.537
	B_5	0.049	-0.065	0.735	0.159	0.749
	B_6	0.204	0.015	0.575	0.386	0.601
	B_7	0.237	0.060	0.609	0.468	0.458
	B_8	0.258	0.041	0.539	0.559	0.273
	C_5	0.044	-0.063	0.738	0.151	0.733
	C_6	0.071	-0.067	0.687	0.222	0.641
	C_7	0.241	0.047	0.570	0.447	0.534
	C_8	0.273	0.088	0.607	0.523	0.405

Table 8. Details of training and testing datasets

Group	Label	Number of the training data	Number of test data	Total
1	Non-knock	550	550	1100
2	Knock	350	350	700
	Total	900	900	1800

ble 9 shows that the average accuracy of SBELM is slightly higher than those of KELM and ELM because the parameters of SBELM are not sensitive to its hyperparameters.

Table 9 reveals that the features of GA-VMD integrated with sample entropy, TDSA, ASD and SBELM have the best accuracy of 98.27%, which is highlighted in red in the table. It is noted that ASD and TDSA have high classification accuracies, whereas MFDFA has poor performance. Even though combining MFDFA with other feature extraction methods can improve the overall precision slightly, MFDFA does not contribute too much to the system accuracy. It also appears that MFDFA is not compatible with GA-VMD because it has the worst accuracy. In summary, Table 9 shows that the integration of SBELM with GA-VMD, sample entropy, ASD and TDSA is an accurate classification method for automatic knock detection.

5 Conclusion

In this paper, a novel intelligence engine knock detection system using multiple feature based SBELM algorithm is successfully developed. GA-VMD is used to filter the unavoidable noises, in which GA is applied to obtain the optimal parameters to enhance

Table 9. Accuracies of various combinations of technologies based on the test dataset

Feature extraction	Signal filtering method	ELM	KELM	SBELM
TDSA	Raw data	93.17%	93.71%	93.50%
	EEMD+sample entropy	94.36%	95.12%	95.23%
	GA-VMD+sample entropy	95.66%	97.72%	97.62%
ASD	Raw data	87.43%	91.76%	91.44%
	EEMD+sample entropy	89.49%	91.87%	92.63%
	GA-VMD+sample entropy	92.84%	92.52%	95.88%
MFDFA	Raw data	72.80%	75.40%	74.65%
	EEMD+sample entropy	70.63%	73.23%	73.56%
	GA-VMD+sample entropy	63.59%	64.78%	63.92%
TDSA+ASD	Raw data	92.41%	93.71%	93.50%
	EEMD+sample entropy	94.47%	95.44%	94.58%
	GA-VMD+sample entropy	96.09%	97.72%	98.27%
TDSA+MFDFA	Raw data	92.84%	94.04%	93.72%
	EEMD+sample entropy	94.69%	95.34%	95.12%
	GA-VMD+sample entropy	95.23%	97.39%	96.86%
ASD+MFDFA	Raw data	94.25%	94.36%	93.72%
	EEMD+sample entropy	91.65%	95.44%	94.37%
	GA-VMD+sample entropy	94.04%	95.88%	96.86%
TDSA+ASD +MFDFA	Raw data	93.82%	93.71%	94.37%
	EEMD+sample entropy	93.39%	95.34%	94.26%
	GA-VMD+sample entropy	95.44%	97.18%	97.40%

the noise reduction ability. When the original time domain signals are decomposed into a series of IMFs, IMFs with sample entropy higher than the mean are selected as sensitive subcomponents for signal reconstruction. Multiple methods, including TDSA, MFDFA and ASD, are applied together to extract features from the denoised signals. The features extracted from the reconstructed signals are then classified by SBELM. The experimental results show that the accuracy of the knock detection system built by SBELM is superior to the accuracies of those built by ELM and KELM. Therefore, the integration of GA-VMD with sample entropy, TDSA, ASD, and SBELM is effective for building automatic engine knock detection systems. Although the proposed method is successfully applied to real engines for engine knock detection, the dataset is recorded from a specific engine model. It will be appealing to apply different engine models to further prove the reliability of the proposed method in future work. Moreover, the training and test data for the proposed system can be expanded to cover more engine speeds, engine loads, air-fuel ratios, fuel octane numbers and engine temperatures to enhance the system generalization. In our current work, the proposed GA-VMD method has the limitation of eliminating non-Gaussian noise under heavy noise disturbances. Non-Gaussian noise always exists in automotive propulsion systems, and usually leads to inconsistencies and divergence of the detection system. Therefore, future work should consider the noise rejection capacity by using correntropy to cope with the issue of non-Gaussian noise.

Acknowledgments

This authors would like to thank the financial support from the University of Macao Distinguished Visiting Scholar Program. This research is funded by the Science and Technology Development Fund, Macau SAR (Nos. 0021/2019/A, 0018/2019/AKP, 0008/2019/AGJ), the Multi-Year Research Grant(No.MYRG2019-00137-FST), the National Natural Science Foundation of China (Nos. 61976172, 12002254) and the Natural Science Basic Research Program of Shaanxi (Nos. 020JQ-013, 2020JM-072). This work is also supported in part by the Macao Youth Scholars Program (No. AM201909).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed Consent Informed consent was not required as no human or animals were involved.

References

1. Arindam Roy Apurba Manna, S.M.: A novel memetic genetic algorithm for solving traveling salesman problem based on multi-parent crossover technique. Decision Making: Applications in Management and Engineering **2**(2), 100–111 (2019)
2. Banerjee, K.S.: Generalized inverse of matrices and its applications. Technometrics **15**(1), 197–197 (1973)
3. Bi, F., Li, X., Lin, J., Bi, X., Ma, T., Yang, X., Tang, D., Shen, P.: Knock detection based on recursive variational mode decomposition and multilevel semi-supervised local fisher discriminant analysis. IEEE Access **7**, 122028–122040 (2019)
4. Bi, F., Ma, T., Wang, X.: Development of a novel knock characteristic detection method for gasoline engines based on wavelet-denoising and EMD decomposition. Mechanical Systems and Signal Processing **117**, 517–536 (2019)
5. Bi, F., Ma, T., Zhang, J., Li, L., Shi, C.: Knock detection in spark ignition engines base on complementary ensemble empirical mode decomposition-hilbert transform. Shock and Vibration **2016**, 9507540 (2016)
6. Borg, J.M., Saikalis, G., Oho, S.: Knock signal analysis using the discrete wavelet transform. Tech. Rep. No. 2006-01-0226, SAE (2006)
7. Chang, J., Kim, M., Min, K.: Detection of misfire and knock in spark ignition engines by wavelet transform of engine block vibration signals. Measurement Science and Technology **13**, 1108–1114 (2002)
8. Chen, F., Chen, X., Yang, Z., Xu, B., Xie, Q., Zhang, H., Ye, Y.: A rolling bearing fault diagnosis method based on VMD-multiscale fractal dimension/energy and optimized support vector machine. Journal of Vibroengineering **18**(6), 3581–3595 (2016)
9. Duin, R.P.: The combining classifier: to train or not to train? In: Object recognition supported by user interaction for service robots, vol. 2, pp. 765–770 (2002)
10. Fu, W., Zhou, J., Zhang, Y.: Fault diagnosis for rolling element bearings with VMD time-frequency analysis and SVM. In: 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), pp. 69–72 (2015)
11. Galloni, E.: Dynamic knock detection and quantification in a spark ignition engine by means of a pressure based method. Energy Conversion and Management **64**, 256–262 (2012)

12. Huang, G.B., Zhou, H., Ding, X., Zhang, R.: Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **42**(2), 513–529 (2012)
13. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of Royal Society A: Mathematical, Physical and Engineering Science **454**, 903–995 (1998)
14. Kang, Y., Wang, C.C., Chang, Y.P.: Gear fault diagnosis by using wavelet neural networks. In: International Symposium on Neural Networks, pp. 580–588 (2007)
15. Kearney, M.J.: Knock signal conditioning using the discrete fourier transform and variable detection window length. Tech. Rep. No. 2007-01-1509, SAE (2007)
16. Li, N., Yang, J., Zhou, R., Wang, Q.: Knock detection in spark ignition engines using a nonlinear wavelet transform of the engine cylinder head vibration signal. Measurement Science and Technology **25**, 115002 (2014)
17. Luo, J., Vong, C.M., Wong, P.K.: Sparse bayesian extreme learning machine for multiclassification. IEEE Transactions on Neural Networks and Learning Systems **25**(4), 836–843 (2014)
18. Mali, P., Sarkar, S., Ghosh, S., Mukhopadhyay, A., Singh, G.: Multifractal detrended fluctuation analysis of particle density fluctuations in high-energy nuclear collisions. Physica A: Statistical Mechanics and its Applications **424**, 25–33 (2015)
19. MohamadAbu-Qudais: Exhaust gas temperature for knock detection and control in spark ignition engine. Energy Conversion and Management **37**(9), 1383–1392 (1996)
20. Ollivier, E., Bellettre, J., Tazerout, M., C.Roy, G.: Detection of knock occurrence in a gas SI engine from a heat transfer analysis. Energy Conversion and Management **47**(7-8), 879–893 (2006)
21. Pan, J., Shu, G., Wei, H.: Interaction of flame propagation and pressure waves during knocking combustion in spark-ignition engines. Combustion Science and Technology **186**(2), 192–209 (2014)
22. Papun Biswas, B.B.P.: A fuzzy goal programming method to solve congestion management problem using genetic algorithm. Decision Making: Applications in Management and Engineering **2**(2), 36–53 (2019)
23. Pincus, S.M.: Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences **88**(6), H2297–H2301 (1991)
24. Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and sample entropy. Journal of Mechanical Science and Technology **278**(6), H2039–H2049 (2000)
25. Sujono, A., Santoso, B., Juwana, W.E.: Sound Vibration Signal Processing For Detection and Identification Detonation (Knock) To Optimize Performance Otto Engine. AIP Conference Proceedings **1717**(1), 030003 (2016)
26. Vong, C.M., Tai, K.I., Pun, C.M., Wong, P.K.: Fast and accurate face detection by sparse bayesian extreme learning machine. Neural Computing and Applications **26**(5), 1149–1156 (2015)
27. Vulli, S., Dunne, J., Potenza, R., Richardson, D., King, P.: Time-frequency analysis of single-point engine-block vibration measurements for multiple excitation-event identification. Journal of Sound and Vibration **321**, 1129–1143 (2009)
28. Wang, X.B., Yang, Z.X., Yan, X.A.: Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery. IEEE/ASME Transactions on Mechatronics **23**(1), 68–79 (2017)
29. Wei, Z., Wang, Y., He, S., Bao, J.: A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. Knowledge-Based Systems **116**, 1–12 (2017)

30. Wong, K.I., Vong, C.M., Wong, P.K., Luo, J.: Sparse bayesian extreme learning machine and its application to biofuel engine performance prediction. Neurocomputing **149**, 397–404 (2015)
31. Wong, P.K., hua Zhong, J., Yang, Z., Vong, C.M.: Sparse bayesian extreme learning committee machine for engine simultaneous fault diagnosis. Neurocomputing **174**, 331–343 (2016)
32. Wong, P.K., Zhong, J.H., Yang, Z.X., Vong, C.M.: A new framework for intelligent simultaneous-fault diagnosis of rotating machinery using pairwise-coupled sparse Bayesian extreme learning committee machine. Proceedings of the Institution of Mechanical Engineers Part-C: Journal of Mechanical Engineering Science **231**(6), 1146–1161 (2016)
33. Xiong, Q., Xu, Y., Peng, Y., Zhang, W., Li, Y., Tang, L.: Low-speed rolling bearing fault diagnosis based on EMD denoising and parameter estimate with alpha stable distribution. Journal of Mechanical Science and Technology **31**(4), 1587–1601 (2017)
34. Yang, J.G., Liu, X.F., Lin, B.: DWT based knock detection and knock intensity judgment for a gasoline engine. Transactions of Chinese Society for Internal Combustion Engines **21**(3), 233–238 (2003)
35. Yao, J., Xiang, Y., Qian, S., Wang, S., Wu, S.: Noise source identification of diesel engine based on variational mode decomposition and robust independent component analysis. Applied Acoustics **116**, 184–194 (2017)
36. Yentes, J.M., Hunt, N., Schmid, K.K., Kaipust, J.P., McGrath, D., Stergiou, N.: The appropriate use of approximate entropy and sample entropy with short data sets. Annals of Biomedical Engineering **41**(2), 349–365 (2013)
37. Yi, C., Lv, Y., Dang, Z.: A fault diagnosis scheme for rolling bearing based on particle swarm optimization in variational mode decomposition. IEEE Transactions on Signal Processing **2016**, 9372691 (2016)
38. Zhang, M., Jiang, Z., Feng, K.: Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump. Mechanical Systems and Signal Processing **93**, 460–493 (2017)
39. Zhang, Z., Tomita, E.: Knocking detection using wavelet instantaneous correlation method. JSAE Review **23**, 443–449 (2002)
40. Zhen, X., Wang, Y., Xu, S., Zhu, Y., Tao, C., Xu, T., Song, M.: The engine knock analysis-an overview. Applied Energy **92**, 628–636 (2012)

Appendix A Result of knock resonant frequencies affected by VMD parameters

Table A1. Result of knock resonant frequencies affected by VMD parameters

	Signal	Raw signal	VMD+sample entropy				
			$K=10$, $\alpha=1500$	$K=10$, $\alpha=5000$	$K=10$, $\alpha=10000$	$K=5$, $\alpha=1500$	$K=20$, $\alpha=1500$
Knock resonant frequencies	f_0	✓	✓	✓	✓	✓	✓
	f_1	✓	✓	✓	✓	✓	✓
	f_2	✓	✓	✓	✓	✓	✓
	f_3	✓	✓	✓	✓	✓	✓
	$2f_0$	✓	✓	✓	✓	✓	✓
	$2f_1$	✓	✓	✓	✓	✓	✓
	$2f_2$	✓	✓	✓	✓	✓	✓
	$2f_3$	✓	✓	✓	✓	✓	✓
	$3f_0$	✓	✓	✓	✓	✓	✓
	$3f_1$						
	$3f_2$	✓	✓	✓	✓	✓	
	$3f_3$	✓	✓	✓	✓	✓	
	$4f_0$	✓	✓		✓		
	$4f_1$	✓	✓	✓	✓		
	$4f_2$						
	$4f_3$	✓	✓		✓	✓	
	$5f_0$	✓	✓			✓	
	$5f_1$						
	$5f_2$	✓	✓			✓	
	$5f_3$	✓	✓			✓	

✓ is used to mark the resonant frequency which appears in the different processed signals.

Appendix B Noise reduction ability of GA-VMD under different IMF selection methods

Table B1. Noise reduction ability of GA-VMD under different IMF selection methods

	Signal	Raw signal	GA-VMD+ Correlation Coefficient	GA-VMD+ Energy Ratio	GA-VMD+ Sample Entropy
Knock resonant frequencies	f_0	✓			✓
	f_1	✓	✓		✓
	f_2	✓			✓
	f_3	✓			✓
	$2f_0$	✓	✓		✓
	$2f_1$	✓			✓
	$2f_2$	✓			✓
	$2f_3$	✓			✓
	$3f_0$	✓	✓		✓
	$3f_1$				
	$3f_2$	✓			✓
	$3f_3$	✓			✓
	$4f_0$	✓			✓
	$4f_1$	✓			✓
	$4f_2$				
	$4f_3$	✓			✓
	$5f_0$	✓			✓
	$5f_1$				
	$5f_2$	✓			✓
	$5f_3$	✓			✓