810. Vibration condition monitoring of planetary gears based on decision level data fusion using Dempster-Shafer theory of evidence

Meghdad Khazaee[1], Hojat Ahmadi[2], Mahmoud Omid[3], Ashkan Moosavian[4], Majid Khazaee[5]
[1, 2, 3, 4] University of Tehran, Department of Mechanical Engineering of Agricultural Machinery, Karaj, Iran
[5] AmirKabir University of Technology, Department of Aerospace Engineering, Tehran, Iran
E-mail: [1]*Khazaee.meghdad@ut.ac.ir*, [2]*hjahmadi@gmail.com*, [3]*omid@ut.ac.ir*, [4]*a.moosavian@ut.ac.ir*, [5]*khazaee101@aut.ac.ir*

(Received 20 February 2012; accepted 14 May 2012)

Abstract. In recent years, due to increasing requirement for reliability of industrial machines, fault diagnosis using data fusion methods has become widely applied. To recognize crucial faults of mechanical systems with high confidence, indubitably decision level fusion techniques are the foremost procedure among other data fusion methods. Therefore, in this paper in order to improve the fault diagnosis accuracy of planetary gearbox, we proposed a representative data fusion approach which exploits Support Vector Machine (*SVM*) and Artificial Neural Network (*ANN*) classifiers and Dempster-Shafer (*D-S*) evidence theory for classifier fusion. We assumed the *SVM* and *ANN* classifiers as fault diagnosis subsystems as well. Then output values of the subsystems were regarded as input values of decision fusion level module. First, vibration signals of a planetary gearbox were captured for four different conditions of gear. Obtained signals were transmitted from time domain to time-frequency domain using wavelet transform. In next step, some statistical features of time-frequency domain signals were extracted which were used as classifiers input. The gained results of every fault diagnosis subsystem were considered as basic probability assignment (*BPA*) of *D-S* evidence theory. Classification accuracy for the *SVM* and *ANN* subsystems was determined as 80.5 % and 74.6 % respectively. Then, by using the D-S theory rules for classifier fusion, ultimate fault diagnosis accuracy was gained as 94.8 %. Results show that proposed method for vibration condition monitoring of planetary gearbox based on *D-S* theory provided a much better accuracy. Furthermore, an increase of more than 14 % accuracy demonstrates the strength of *D-S* theory method in decision fusion level fault diagnosis.

Keywords: vibration condition monitoring, Dempster-Shafer theory, data fusion, support vector machine, artificial neural network.

Introduction

Accordingly, reliability and accessibility of equipment always has great importance [1]. With these advances in technology, industrial machinery increasingly became complicated and more sensitive simultaneously that require further attention. Their fault and failure might lead to overwhelming costs [2].Vibration condition monitoring has long been accepted as one of the most effective approaches for fault diagnosis and classification of machines [3]. Many methods were presented and conducted based on vibration signals [4, 5] by using one classifier like support vector machine [6], artificial neural network [7] or fuzzy logic [8]. The studies indicate that using only one classifier might produce many limitations. A single classifier cannot get the fault diagnosis accuracy precisely and has limited classification capability which might not be sufficient for a diagnostic application [2, 9, 10]. This aspect emphasizes the importance of using data fusion in condition monitoring applications.

Recently, data fusion and its applications have attracted significant attention and are widely applied in many areas such as robotics [11], artificial intelligence [12], condition monitoring and fault diagnosis [2, 9], pattern recognition [13], etc. Data fusion techniques combine data and related information to achieve improved accuracies and recognition precision, reducing the uncertainty [14]. In data fusion systems, data may be fused at three levels, namely, Signal (data) level fusion, Feature fusion level and Decision level fusion [14].

Recently, the decision fusion (or classifier fusion) based on multi-expert combination strategies, has been of major concern [15] and researchers have reached noticeable successes from this approach to make complex pattern recognition tasks glaringly obvious [2]. Classifier fusion methods can give more dependable results in comparison with any individual classifier [16]. But the question is to how fuse the classifiers. Many methods have been applied to combine the classifiers such as Bayesian inference [17], Dempster-Shafer theory [18, 19] and etc. To obtain more details see [20-23]. Dempster-Shafer (*D-S*) theory of evidence is one of the most common and powerful methods for fusing the classifiers results, which attracted considerable attention and many research studies have been performed by using *D-S* theory for fault diagnosis of machines. Niu and Yang [24] present a method based on *D-S* theory for machinery prognosis. Yang and Kim [18] combined vibration signals and current signals at decision level using *D-S* theory for fault diagnosis of induction motors. Also see [25] and [26].

In this paper we propose a decision fusion method based on Support Vector Machine (*SVM*) and Artificial Neural Network (*ANN*) as local classifiers and *D-S* theory of evidence to fuse results in order to improve precision of vibration fault diagnosis and classification of planetary gearbox. *SVM* and *ANN* were assumed as classification subsystems. The extracted features from wavelet decomposition of vibration signals were used to feed subsystems. The output values of *SVM* and *ANN* classifiers are regarded as input value of decision fusion method. Local fault diagnosis accuracies get as basic probability assignment (*BPA*) of *D–S* theory. Then by using the *D-S* theory rules, the results were fused and the final accuracy was determined.

Planetary gearbox is among the most conventional gearboxes widely considered for their wide range of gear ratios and in particular in helicopters and heavy machineries. Low weight and small room in power transmission lines are their most significant advantages. These gearboxes are generally composed of three parts: 1 - sun gear, 2 - ring gear, 3 - planet gears. Fig. 1 displays the general structure of planetary gearbox and the relationship between its components. In this study, one of the varieties of planetary gears, namely, MF285 tractor final drive was used. Final drive is the last component of power transmission system in all tractors which is located between differentials and wheels and its function is to increase torque on the wheels and decrease the speed.

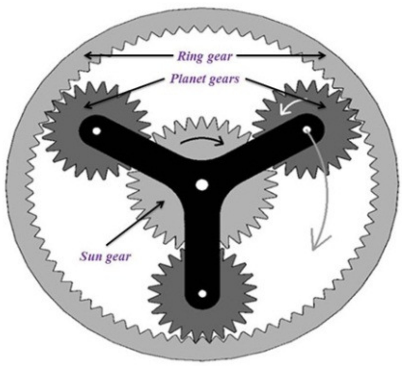

Fig. 1. Planetary gearbox and its components

Experimental setup

The experimental setup to collect dataset consists of MF285 final drive, an electromotor that assembles drive power using a coupling. A test-bed was built to mount gearbox, electromotor, and other equipment. The planetary gearbox coupled to the electromotor that was initially run under normal operating conditions and its speed at 300 RPM was controlled by an inverter. Fig. 2 illustrates the experimental setup that was used in this study.

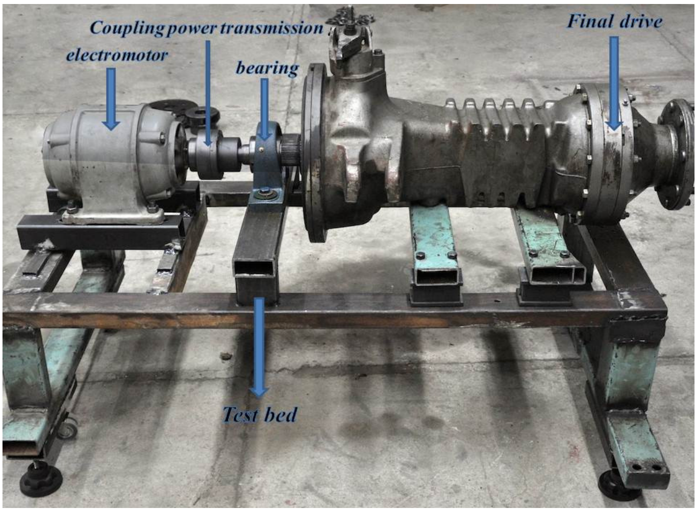

Fig. 2. Experimental setup included final drive, test bed and etc.

The faults were manually created on gears. Three common gears fault, namely, 1) broken ring gear, 2) cracked ring gear, 3) planet gear with worn tooth and healthy state were the four conditions that were studied in this research.

Common techniques used for vibration fault detection of gears include time domain, frequency domain and time-frequency domain analyses. In this paper, the vibration behaviors of the planetary gears for normal and defective gears were studied. The velocity vectors (mm/sec) of gearbox vibration were measured and assumed as vibration signals. The root mean square (*RMS*) of vibration velocity was calculated for determination the time domain vibration signals. Then, in the signal processing step, time domain signals were transmitted to time-frequency domain by Wavelet transform. Thereby, it is possible to collect time and frequency data of any position in a signal and have better performance in extracting the features and classification of the faults. This technique for vibration signals analysis is an effective and efficient method for fault diagnosis of planetary gears. For more details see some similar studies in [27- 29].

In data collection step, the sensor was located horizontally on the surface of final drive and vibration signals of each of the classes were captured from the experimental testing with an accelerometer of type VMI102.

Signal processing

Although vibration signals carry useful data about the machine conditions, they include not only fault signals but they carry all the component signals and many noise signals that may aggravate fault detection [30]. In order to solve this problem the time domain signals must be transferred to the frequency domain or time-frequency domain. As a result, this action removes noises, as it allows gaining more useful data. Various techniques have been applied for signal processing such as Wavelet analysis, Fast Fourier Transform (*FFT*), Short Time Fourier Transform (*STFT*), etc.

In present study, Wavelet transform was used to transform data from the time domain to time-frequency domain. Continues wavelet transform (*CWT*) decomposes signal in both time and frequency domains simultaneously. *CWT* is defined as:

$$CWT(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^{*}\left(\frac{t-\tau}{a}\right) dt \tag{1}$$

where a, τ and ψ and are the scale parameter, translation parameter and mother wavelet, respectively, and ψ^{*} is the complex conjugate of ψ. Computation of wavelet coefficients at every conceivable scale takes lots of calculations and work, and result in huge and awful amount of data [31]. Thus, using dyadic scale and positions, makes analysis more efficient and accurate, this procedure is called discrete wavelet transform (*DWT*). *DWT* which is the discrete form of CWT is expressed as:

$$DWT(a,\tau) = \frac{1}{\sqrt{2^{j}}} \int_{-\infty}^{+\infty} x(t)\psi^{*}\left(\frac{t-2^{j}k}{2^{j}}\right) dt \tag{2}$$

where a and τ represent 2^j and 2^jk. The *DWT* analysis is done by passing the raw signal through a series of high and low pass frequency filters. Each level of decomposition consist of one high-pass and one lower-pass filter, thus, the raw signal is decomposed into two parts, high frequency bands (Details (D_j)) and low frequency bands (Approximation (A_j)). In next level of decomposition, approximation signal of previous level is applied as input of decomposition and high and low frequency bands will be separated, and this process is done till reach of desired decomposition level. The original signal can be defined as [7]:

$$x(t) = A_j + \sum_{j \le J} D_j \tag{3}$$

where A_j and D_j are approximation and detail of the signal at level J^{th}, respectively. In other words, this signal is the composition of J^{th} level details and last level approximations wavelet coefficients. Processing of the signals was conducted using wavelet toolbox in MATLAB.

Local classification: *SVM* and *ANN*

In current study, two subsystems were defined to perform local classification. Two classifiers were employed through local classification of vibration signals, namely, Support Vector Machine (*SVM*) and Artificial Neural Network (*ANN*).

1. Support Vector Machine (SVM)

Support vector machine (*SVM*) is a new computational supervised learning method for classification which has been developed by Vapnik and is gaining popularity due to many appealing features. *SVM* is based on *VC*-theory and attempts to construct, and then seek the separating hyper planes with maximum margin by converting the problem description into dual space [32]. This is done by the means of Lagrangian and solving the dual optimization problem. The *SVM* is introduced based on structural risk minimization (*SRM*), which minimizes the upper bound of the generalization error. It gives better generalization abilities than methods used empirical risk minimization (*ERM*) such as Neural Network [32, 33]. *SVM* have been successfully applied to the number of applications such as machine condition monitoring, fault diagnosis, face detection, verification, and recognition, object detection and recognition, handwritten character and digit recognition, text detection and categorization, speech and speaker verification, recognition, information and image retrieval, etc.

SVM has the potential to handle huge feature spaces [34]. Support vector machines (*SVM*) were originally designed for binary (2-class) classification. In binary classification, the class labels can take only two values: 1 and -1. In the case of nonlinear classification, *SVM* uses the kernel function. The kernel functions provide a solution to this problem by adding an additional dimension to the dataset. In other words, a kernel function maps data from a low-dimensional space (original space) onto a space of higher dimension, where the linear classification is possible [35]. Using the nonlinear function, the *N*-dimensional input vector x_i is mapped onto the *l*-dimensional feature space. Kernel is a function that returns a dot product of the feature space mappings of the original vectors. In this case, knowing explicitly the mapping function is not required for the learning.

2. Artificial Neural Network (ANN)

Artificial neural network (*ANN*) is one of the most widely used artificial intelligence methods in condition monitoring and fault diagnosis [7] and also it was used for classification the different conditions of gears in this study. Outlet layer of all networks included four neurons, since it was defined as a 1-row-4- column matrix with 0-1 digits to define desired class. This means 0-1-0-0 and 0-0-1-0 outputs respectively is related to second and third classes of the quartet classes.

Nevertheless, the most important layer in designing neural network is the hidden layer (middle layer) that should be defined by settings. In current research, networks with 'tansig' transfer function and 'mse' performance function and variable count, between 2 and 10, are used to gain the best classification results. To meet the best network performance, the number of middle layers were tested among which the best results gained with 10 neurons of the middle layer. In designing, a neuron was assigned to each extracted feature and since there were 30 features extracted, for the sake of one sensory fault diagnosis, network input had 30 neurons. Finally, regarding the explanations, optimal neural network structure designed in this study was 30*10*4.

Dempster-Shafer theory of evidence

D-S theory of evidence was produced by Shafer in 1976 [36] as a powerful tool for representing uncertain knowledge. This theory has inspired many researchers to investigate different aspects connected with uncertainty and lack of knowledge and their applications in real life problem [16].

1. Basic functions of D-S theory

If define $\Omega = \{\Theta_1, \Theta_2, \ldots, \Theta_k\}$ which Ω is a finite set of possible hypotheses, this set is referred to the frame of discernment and the power set of set denote by 2^Ω [16]. There are three basic functions in *D-S* theory: the Basic Probability Assignment function (*BPA* or *m*), the Belief function (*Bel*), and the Plausibility function (*Pl*) [36].

The basic probability assignment (*BPA*) is a primitive of evidence theory. This function assigns a value in [0, 1] to set *A*. The value of *m*(*A*) pertains only to the set *A* and makes no additional claims about any subsets of *A* [36, 37]. *PBA* can be described by the following three equations:

$$\begin{cases} m : P(X) \to [0,1] \\ \text{and} \\ m(\phi) = 0 \\ \sum_{A \in P} m(A) = 1 \end{cases} \tag{4}$$

where $P(X)$ represents the power set of X, ϕ is the null set, and A is a set in the power set $A \in P(X)$ [37]. Two other functions called belief function (*Bel*) and Plausibility function (*Pl*). *Bel* and *Pl* are the upper and lower bounds of an interval respectively and can be defined from the basic *BPA* function. The *Bel*(*A*) (lower bound) is defined as sum of all the basic probability assignments of the proper subsets (*B*) which $B \subseteq A$. The *Pl* (upper bound) is the sum of all the basic probability assignments of the sets (*B*) that $B \cap A \neq \phi$ [37]. *Bel*(*A*) and *Pl*(*A*) can be represented with the following equations:

$$Bel(A) = \sum_{B|B \subseteq A} m(B) \tag{5}$$

$$Pl(A) = \sum_{B|B \cap A \neq \phi} m(B) \tag{6}$$

By attention to (5) and (6) it is apparent that $Bel(A) \leq Pl(A)$. Finally note that relevancy between *Bel* function and *Pl* function:

$$Pl(A) = 1 - Bel(\overline{A}) \tag{7}$$

Precision probability *P*(*A*) of event A, lies within of *Bel* and *Pl,* respectively [38]:

$$Bel(A) \leq P(A) \leq Pl(A) \xrightarrow{If \;\; Pl(A) = Bel(A)} P(A) = Pl(A) = Bel(A) \tag{8}$$

2. D-S theory rules for classifier combination

The combination rules are critical to the original conception of Dempster-Shafer theory. The measures of Belief and Plausibility are derived from the combined basic assignments. The Dempster rule of combination is purely a conjunctive operation (*AND*). The combination rule results in a belief function based on conjunctive pooled evidence [39]. The combination rules of two *BPA* and multi *BPA* are presented in (9) and (10) respectively [38]:

$$m_{12}(A) = m_1(B) \oplus m_2(C) = \frac{\sum_{B \cap C = A} m_1(B) \times m_2(C)}{1 - \sum_{B \cap C = \phi} m_1(B) \times m_2(C)}, \quad \text{when} \quad A \neq \phi \tag{9}$$

$$\begin{cases} m_1(A) \oplus m_2(A) \oplus \ldots \oplus m_i(A) = (1-K)^{-1} \times \sum_{\bigcap A_i = A} m_i(A_i) \\ K = \sum_{\bigcap A_i = \phi} m_i(A_i) \end{cases} \tag{10}$$

where K represents basic probability mass associated with conflict. This is determined by summation of all *BPA* product sets while intersection is null. This rule is commutative, associative, but not idempotent or continuous [40].

3. Yager's rule for combination

The denominator in *D-S* theory rule, (1-*K*), is a normalization factor. This factor has effect of completely ignoring conflict and attributing any probability mass associated with the conflict to the null set. Yager improved *D–S* evidence theory by classifying conflicting evidences into set Θ [38]. Θ means that the classifier does not know the class of input data. In Yager's rule Θ and $q(A)$ for two evidences can present by following equations [38]:

$$\Theta_i = 1 - \alpha_i \tag{11}$$

$$q(A) = \sum_{\bigcap A_i = A} m_1(A_i).m_2(A_i) + \sum_{\bigcap A_i = A} m_1(A_i) \times \Theta_2 + \sum_{\bigcap A_i = A} m_2(A_i) \times \Theta_1 \tag{12}$$

where α_i is the weight of evidences (also called importance factor of evidence). As previously mentioned, one obvious distinction between combination with the basic and the ground probability assignment functions is the absence of the normalization factor (*1-K*). In Yager's formulation, he circumvents normalization by allowing the ground probability mass assignment of the null set to be greater than 0, so [38]:

$$q(\phi) \geq 0 \tag{13}$$

The symbol ϕ in (14) means that the conflict occurred between two classifiers [18]. Many research studies have been done with aim to determine α_i as a fundamental factor of combination rule. For example, Yang and Kim [18] determined these coefficients by training data set for each evidences. For more details see [26]. In this study we assumed train accuracy of any subsystem in title of that subsystem weight. Now the new combination rule for two evidences can be presented with the following function [38]:

$$m(A) = \frac{q(A)}{1 - q(\phi)} \tag{14}$$

Experiments and results

In this work, first, in data collecting step, vibration signals were gathered in four different conditions of gears. Each one of the four classes had 60 samples and, in sum, 240 samples. Each class of data was divided into two parts: 45 samples (75 %) out of each 60 samples of each class were selected entirely random and applied to train the *ANN* and *SVM* classifiers and other 15 samples were used for testing the system. Fig. 3 shows the kind of classes that was classified in this study. Gained signals were transferred from time domain to time-frequency domain by wavelet transform. Fig. 4 shows vibration signals in time domain. Also, wavelet decomposition of vibration signals is shown in Fig. 5.

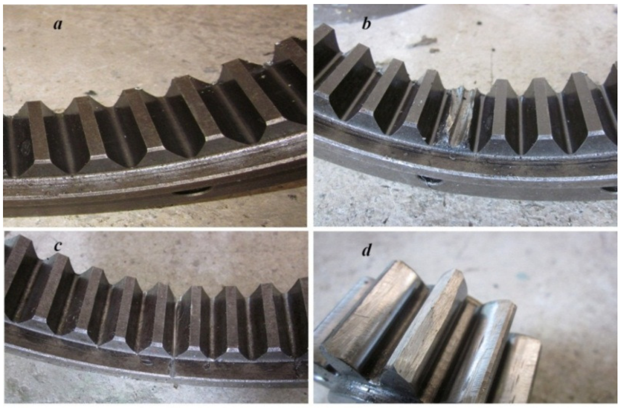

Fig. 3. Different conditions of gears:
a) intact, b) broken ring gear, c) cracked ring gear, d) planet gear with worn tooth face

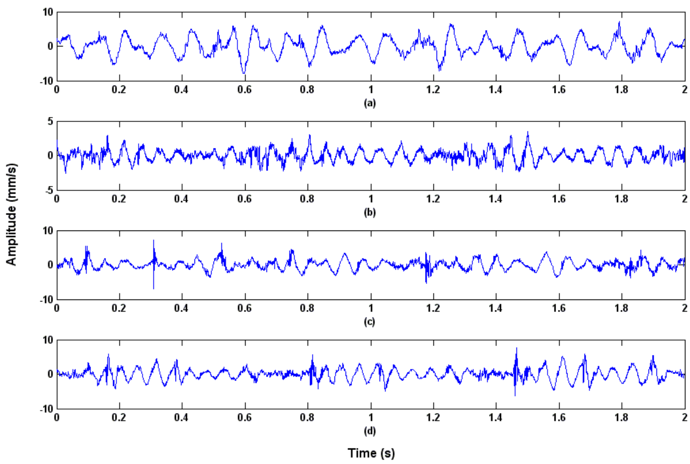

Fig. 4. Time domain vibration signals:
a) intact, b) planet gear with worn tooth face, c) broken ring gear, d) cracked ring gear

Processed signals contain a large set of data for each sample therefore we applied some functions to reduce feature vectors. In current study 30 features were extracted from the wavelet coefficients in approximation and details such as mean, standard deviation, kurtosis, skewness, etc. The extracted features were employed to feed the subsystems to determine local fault classification accuracy. Some important features and their formulas are shown in Table 1.

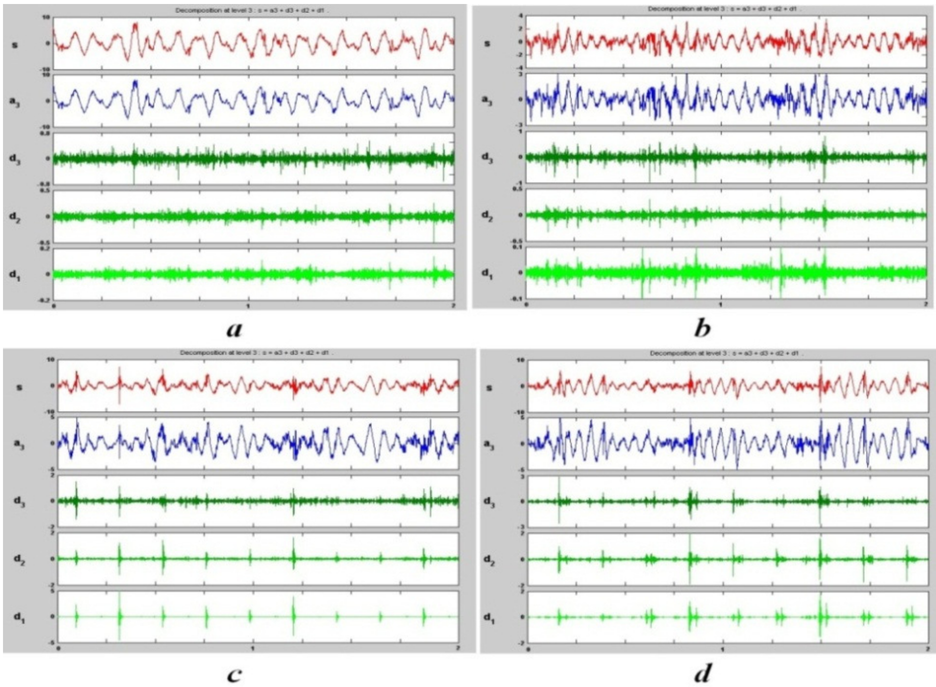

Fig. 5. Wavelet decompositions of vibration signals:
a) intact, b) planet gear with worn tooth, c) broken ring gear, d) cracked ring gear

Table 1. Some features and their formulas that used to feed classifier

Feature Description	Formula	Feature Description	Formula
Mean value	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	Third central moment	$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{n}$
Standard deviation	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$	Forth central moment	$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{n}$
Root mean square	$\sqrt{\frac{\sum_{i=1}^{n} (x_i)^2}{n}}$	Kurtosis	$\frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2)^2} - 3$
Skewness	$\frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^3}{(\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2)^{3/2}}$	FM4	$\frac{n\sum_{i=1}^{n} (x_i - \bar{x})^4}{(\sum_{i=1}^{n} (x_i - \bar{x})^2)^2}$

1. First evidence: SVM classifier

In subsystem 1, extracted features from vibration signals were applied to determine the classification accuracy of *SVM* classifier. As accuracy of the *SVM* classifier for training data was gained as 94 %, weight of the evidence 1 was assumed as $\alpha_1 = 0.94$. Let $O_i(F_j)$ is the output accuracy of i^{th} subsystem for F_j fault. Then, final BPA for i^{th} evidence and F_j fault is:

$$m_i(F_j) = \alpha_i \times O_i(F_j) \tag{15}$$

Also, $(1-\alpha_i)$ means the belief mass assigned to the initial hypothesis set Θ_i [18]. Therefore, Θ_1 was determined as 0.06 for the first evidence. Table 2 shows accuracy results of *SVM* classifier. These values were assumed as *BPA* for the first evidence in *D-S* theory rules.

Suppose that: [Healthy; Broken Ring; Cracked Ring; Worn Planet] = [H; BR; CR; WP].

Table 2. *BPA*'s of subsystem1: *SVM* classifier

classes	*H*	*BR*	*CR*	*WP*	Θ_1
H	0.83	0.016	0.0329	0.0611	0.06
BR	0.0103	0.8601	0.0517	0.0179	
CR	0.0235	0.0517	0.7586	0.1062	
WP	0.0254	0.0498	0.0912	0.7736	

2. Second evidence: ANN classifier

In this subsystem, the extracted features were used to feed *ANN* classifier too. *ANN* Accuracy for training data gained as 89 %. So the importance factor of the second evidence was assumed as $\alpha_2 = 0.89$ and Θ_2 determined as 0.11. *BPA*'s values of the second evidence are shown in Table 2.

Table 3. *BPA*'s of subsystem2: ANN classifier

classes	H	BR	CR	WP	Θ_1
H	0.7618	0.0231	0.0418	0.0632	0.11
BR	0.0142	0.8081	0.0285	0.0392	
CR	0.0258	0.0329	0.7156	0.1157	
WP	0.0498	0.0339	0.1068	0.6995	

3. Combination of classifiers results

In this step, *Dempster-Shafer* rules for classifier fusion were applied to combine the output *BPA*'s of *SVM* and *ANN* classifiers in order to increase overall accuracy of fault diagnosis method. In this paper we present a decision-level fusion aim to improve the classification performance. Fig. 3 presents architecture of the proposed method in this article for decision level fusion.

For combination the *BPA*'s we used (12) and (14) equations:

$$q(A) = \sum_{\bigcap A_i = A} m_1(A_i).m_2(A_i) + \sum_{\bigcap A_i = A} m_1(A_i)\times\Theta_2 + \sum_{\bigcap A_i = A} m_2(A_i)\times\Theta_1 \tag{12}$$

$$m(A) = \frac{q(A)}{1 - q(\phi)} \tag{14}$$

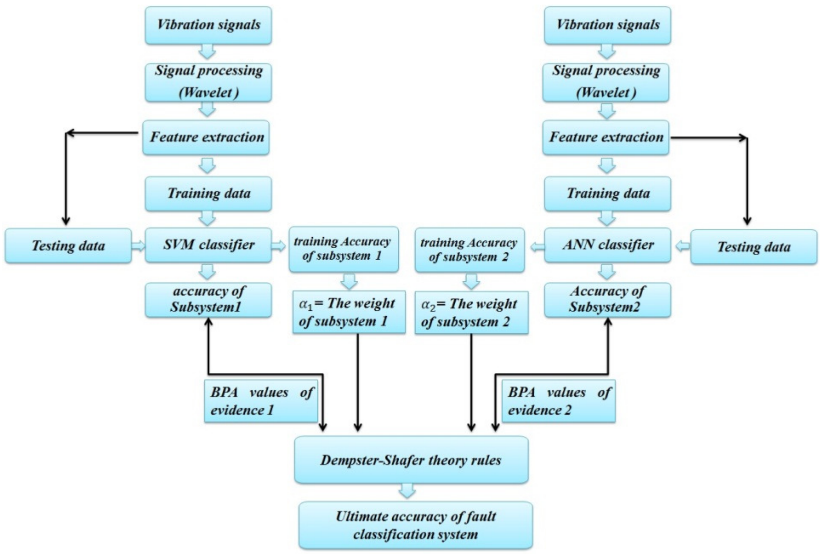

Fig. 6. The structure of the proposed method for classifier combination

For example, the combination process of *ANN* and *SVM* classifier's *BPA* for "Broken ring" fault is shown in Table 4. Also, Table 5 represents calculation details to determine combined BPA for this state of gears.

Table 4. Combination process of *SVM* and *ANN* classifiers for '*BR*' gear condition

	Classification results of *SVM* classifier				
Classification results of *ANN* classifier	*H* 0.0103	*BR* 0.8601	*CR* 0.0517	*WP* 0.0179	Θ_2 .06
H 0.0142	0.0001	0.0122 ϕ	0.0007 ϕ	0.0003 ϕ	0.0009
BR 0.8081	0.0083 ϕ	0.695	0.0418 ϕ	0.0145 ϕ	0.0485
CR 0.0285	0.0003 ϕ	0.0245 ϕ	0.0015	0.0005 ϕ	0.0017
WP 0.0392	0.0004 ϕ	0.0337 ϕ	0.002 ϕ	0.0007	0.0024
$\Theta_2 = 0.11$	0.0011	0.0946	0.0057	0.002	0.0066

Table 5. Detail calculation of combination process for '*BR*' state

Gear conditions	$q(A)$	$q(\phi)=\sum \phi_i$	$q(A)/(1-q(\phi))$	$m(A)$
H	0.0021	0.1392	0.0021/(1-0.1392)	0.0025
BR	q=0.695+0.0964+0.0485=0.831		0.831/(1-0.1392)	0.9737
CR	0.0089		0.0089/(1-0.1392)	0.0103
WP	0.0051		0.0051/(1-0.1392)	0.0058

The combined classification accuracy for each condition of gears is shown in Table 6. Table 7 shows all the results of this research. In addition, it illustrates a comparison among the classification accuracy. As this table implies, average classification accuracy has been increased about 15 % after classifiers results fusion by *D-S* theory.

Table 6. Fused *BPA* of two classifiers

Gears condition	*H*	*BR*	*CR*	*WP*
H	0.96	0.0044	0.0094	0.0179
BR	0.0025	0.9737	0.0103	0.0058
CR	0.0066	0.013	0.9284	0.0429
WP	0.0098	0.0128	0.0365	0.9316

Table 7. Final accuracy and classifiers results comparison

Classifier	Classification accuracy				
	H	*BR*	*CR*	*WP*	Average
SVM	0.83	0.8601	0.7586	0.7736	0.8055
ANN	0.7618	0.8081	0.7156	0.6995	0.7462
Combined results	0.96	0.9737	0.9284	0.9316	0.9484

The results obviously illustrates that application of data fusion has increased accuracy satisfactorily. Also Dempster-Shafer theory was successful at combining two separate classifiers and displayed its capacity for fusing uncertain data.

Conclusions

In this article, we presented a powerful method for fault diagnosis and classification of planetary gearbox based on gear vibration signals. With the aim to have a better feature extraction, vibration signals were transformed from the time domain to the time-frequency domain by means of Wavelet transform. Then, 30 features were extracted from wavelet decomposition of vibration signals to feed classifiers. *SVM* and *ANN* classifiers were applied as local classifiers. The output of these classifiers was assumed as inputs of Dempster-Shafer theory in order to improve classification accuracy. As it was shown in Table 7, average accuracy for the *SVM* and *ANN* classifiers were gained as 80.55 % and 74.62 %, respectively. Combination of two classifier using *D-S* theory rules led to ultimate accuracy of 94.84 %. This increase of about 14 % indicates efficiency of *D-S* theory for decision fusion of classifier results. Also, the results demonstrate high capability and appropriate quality of the introduced system in this article for vibration condition monitoring and fault diagnosis of planetary gearbox.

References

[1] **Jardine A. K. A., Lin D., Banjevic D.** A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, Vol. 20, 2006, p. 1483–1510.

[2] **Niu G., Han T., Yang B. S., Tan A. C. C.** Multi-agent decision fusion for motor fault diagnosis. Mechanical Systems and Signal Processing, Vol. 21, 2007, p. 1285–1299.

[3] **Mollazade K., Ahmadi H., Omid M., Alimardani R.** Vibration-based fault diagnosis of hydraulic pump of tractor steering system by using energy technique. Modern Applied Science, Vol. 3, Issue 6, 2009, p. 5966.

[4] **Wang X., Makis V., Yang M.** A wavelet approach to fault diagnosis of a gearbox under varying load conditions. Journal of Sound and Vibration, Vol. 329, 2010, p. 1570–1585.

[5] **Miao Q., Makis V.** Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models. Mechanical Systems and Signal Processing, Vol. 21, 2007, p. 840–855.

[6] **Yuan S., Fulei C.** Fault diagnosis based on support vector machines with parameter optimization by artificial immunization algorithm. Mechanical Systems and Signal Processing, Vol. 21, 2007, p. 1318–1330.

[7] **Bagheri B., Ahmadi H., Labbafi R.** Implementing discrete wavelet transform and artificial neural networks for acoustic condition monitoring of gearbox. Elixir Mech. Engg., Vol. 35, 2011, p. 2909–2911.

[8] **Mollazade K., Ahmadi H., Omid M., Alimardani R.** An intelligent combined method based on power spectral density, decision trees and fuzzy logic for hydraulic pumps fault diagnosis. World Academy of Science, Engineering and Technology, Vol. 44, 2008, p. 895907.

[9] **Dengchao F., Dias Pereira J. M.** Study on information fusion based on wavelet neural network and evidence theory in fault diagnosis. The Eighth International Conference on Electronic Measurement and Instruments, Xian, China, August 2007, p. 35223526.

[10] **Zhang X., Tian F., Mu Y.** Neural network integration fusion model and application. Third International Conference on Natural Computation (ICNC2007), Hainan, China, August 2007, p. 13061309.

[11] **Abidi M. A., Gonzalez R. C.** Data Fusion in Robotics and Machine Intelligence. Boston: MA Academic, 1992.

[12] **Hall D. L., Linn R. J.** Survey of commercial software for multi-sensor data fusion. In Proc. SPIE Conf. Sensor Fusion and Aerospace Applications, Orlando, 1991.

[13] **Gunatilaka A. H., Baertlein B. A.** Feature-level and decision-level fusion of non-coincidently sampled sensors for land mine detection. IEEE Trans. Pattern Anal. Mach. Intell., Vol. 23, Issue 6, 2001, p. 577–589.

[14] **Hall D. L., Llinas J.** An introduction to multisensor data fusion. Proceedings of the IEEE, Vol. 85, Issue 1, 1997, p. 623.

[15] **Yang J., Yang J. Y., Zhang D., Lu J. F.** Feature fusion: parallel strategy vs. serial strategy. Pattern Recognition, Vol. 36, 2003, p. 13691381.

[16] **Alani A., Deriche M.** A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence. Journal of Artificial Intelligence Research, Vol. 17, 2002, p. 333361.

[17] **Iverson G. R.** Bayesian Statistical Inference. Beverly Hills: CA Sage, 1984.

[18] **Yang B. S., Kim K. J.** Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals. Mechanical Systems and Signal Processing, Vol. 20, 2006, p. 403–420.

[19] **Eggers M., Khuon T.** Neural network data fusion for decision making. In Proc. 1989 Tri-Service Data Fusion Symp., May 1989, p. 104–118.

[20] **Dai X., Khorram S.** Data fusion using artificial neural networks: a case study on multi temporal change analysis. Computers, Environment and Urban Systems, Vol. 23, 1999, p. 1931.

[21] **Hall D.** Mathematical Techniques in Multisensor Data Fusion. Boston: Artech House, 1992.

[22] **Kittler J., Hatef M., Duin R., Matas J.** On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, 1998, p. 226–239.

[23] **Ho T., Hull J., Srihari S.** Decision combination in multiple classifier system. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, 1994, p. 66–75.

[24] **Niu G., Yang B. S.** Dempster-Shafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis. Mechanical Systems and Signal Processing, Vol. 23, 2009, p. 740–751.

[25] **Basir O., Yuan X.** Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory. Information Fusion, Vol. 8, 2007, p. 379–386.

[26] **Fan X., Zuo M. J.** Fault diagnosis of machines based on D–S evidence theory. Part 1: D–S evidence theory and its improvement. Pattern Recognition Letters, Vol. 27, 2006, p. 366–376.

[27] **Blunt D. M., Keller J. A.** Detection of a fatigue crack in a UH-60A planet gear carrier using vibration analysis. Mechanical Systems and Signal Processing, Vol. 20, 2006, p. 20952011.

[28] **Bartelmus W., Zimroz R.** Vibration condition monitoring of planetary gearbox under varying external load. Mechanical Systems and Signal Processing, Vol. 23, 2009, p. 246–257.

[29] **McFadden P. D.** A technique for calculating the time domain averages of the vibration of individual planet gears and sun gear in an epicyclic gearbox. Journal of Sound and Vibration, Vol. 144, 1991, p. 163–172.

[30] **Kunpeng Z., San W. Y., Soon H. G.** Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results. Expert Systems with Applications, Vol. 49, 2009, p. 537–553.

[31] **Wu J., Chan J.** Faulted gear identification of rotating machinery based on wavelet transform and artificial neural network. Expert Systems with Applications, Vol. 36, 2009, p. 88628875.

[32] **Vapnik V. N.** The Nature of Statistical Learning Theory. New York: Springer, 1995.

[33] **Cortes C., Vapnik V.** Support-vector networks. Machine Learning, Vol. 20, Issue 3, 1995, p. 273–297.

[34] **Widodo A., Yang B. S.** Wavelet support vector machine for induction machine fault diagnosis based on transient current signal. Expert Systems with Applications, Vol. 35, 2008, p. 307–316.

[35] **Noble W. S.** What is support vector machine?. Nature Biotechnology, Vol. 24, Issue 12, 2006, p. 15651567.

[36] **Shafer G.** A Mathematical Theory of Evidence. Princeton: Princeton Univ. Press, 1976.

[37] **Klir G. J., Wierman M. J.** Uncertainty-Based Information: Elements of Generalized Information Theory. Heidelberg: Physica-Verlag, 1998.

[38] **Yager R.** On the Dempster-Shafer framework and new combination rules. Information Sciences, Vol. 41, 1987, p. 93–137.

[39] **Shafer G.** Probability Judgement in Artificial Intelligence. Uncertainty in Artificial Intelligence. New York: Elsevier Science, 1986.

[40] **Sentz K., Freeson S.** Combination of Evidence in Dempster-Shafer Theory. California: Sandia National Laboratories, 2002.