International Journal of Artificial Intelligence and Applications for Smart Devices
Vol.1, No.1 (2013), pp.41-64
http://dx.doi.org/10.14257/ijaiasd.2013.1.1.04

Fault Gear Categorization: A Comparative Study on Feature Classification using Rough Set Theory and ID3

Saravanan Natarajan

Department of Engineering- Mechanical and Industrial Engineering, Higher College of Technology, Ministry of Manpower, Muscat, Sultanate of Oman

Saravanan.natarajan@hct.edu.om, natarajan.saravanan1@gmail.com

Abstract

Fault diagnosis on a gear box is a difficult problem due to the non-stationary type of vibration signals it generates. Usually, one method of fault diagnosis can only inspect one corresponding fault category. Vibration based condition monitoring using machine learning methods is gaining momentum. In this paper, rough sets theory, is used to diagnose the fault gears in a gear box. Through the analysis of the final reducts generated using rough sets theory, it is shown that this method is effective for diagnosing more than one type of fault in a gear. The performance of rough set method are compared with those of the ID3 decision tree algorithm and the results prove that the rough set method has greater capability to bring out the different fault conditions of the gear box under investigation. The study reveals that the overall classification efficiency of the decision tree is to some extent better than the classification efficiency of rough sets method.

Keywords: *Rough set method, Bevel gear box, ID3, Fault classification, Statistical features, Fault detection*

1. Introduction

A faulty gear system could result in serious damage if defects occur to one of the gears during operation. Early detection of the defects, therefore, is crucial to prevent the system from malfunctioning, which could cause damage or entire system halt. Diagnosing a gear system by examining vibration signals is the most commonly used method for detecting gear failures. The conventional methods for processing measured data include the frequency domain technique, time domain technique, and time-frequency domain technique. These methods have been widely employed to detect gear failures. The use of vibration analysis for gear fault diagnosis and monitoring has been widely investigated and its application in industry is well established [1-3]. However, up to now, it has been difficult to diagnose more than one category of faults using conventional methods. This is especially so in diagnosing the dynamic characteristics of rotating machinery, such as compressors, gears and engines. This is due to the complex structure of the machinery and the nature of vibration signals generated [4].There has to be a method to diagnose more than one category of faults in a generic manner. A method based on rough sets theory is proposed and implemented in this paper, Z. Pawlak (Poland) first proposed rough sets theory in 1982. This method has been used in many areas, such as medical diagnosis [5], stock market forecasting [6], fault diagnosis in the engineering domain [7], decision making for banking [8] and some other uses [9]. Rough set theory (RST), as proposed by Pawlak, is a useful mathematical approach to solve the problem with imprecise, uncertain and vague information. Owing to its advantage that the importance of various attributes is evaluated and certain key attributes are retained

ISSN: 2288-6710 IJAIASD

with no additional knowledge except for the supplied data required, RST, as an important information processing tool, has achieved excellent performance in a wide variety of applications including data mining, pattern recognition and fault diagnosis [11-15].

Recent research progress in database technologies has created a significant interest in knowledge discovery in databases and data mining [16]. Knowledge discovery refers to the automation of knowledge extraction from large databases [17-19]. A wide variety of artificial intelligence techniques are used for rule induction from these large databases [20], and algorithms are developed to learn the regularities from the rich data [21]. These techniques include neural networks [22], ID3 (Decision tree) [23, 24] and rough sets [25].

RST and ID3 eliminate superfluous attributes in the process of determining significant attributes for classification. It has been shown that, under special circumstances, when the distribution of objects in the boundary region of a rough set is equally probable, the criteria for selecting dominant attributes is a special case of ID3 [26]. A large amount of literature exists for the classification of faults in gear boxes, but the comparison of the RST and ID3 rules have not been reported widely.

In this paper, RST and the ID3 algorithm are applied to diagnose the fault types simulated in a gear box. The vibration signals are extracted from the different conditions of the gear box. Then the statistical features of the vibration data are extracted and classified using ID3 rules and RST. Gear boxes with the following conditions of the gear box are studied: good bevel gear, bevel gear with tooth breakage (GTB), bevel gear with crack at root of the tooth (GTC), and bevel gear with face wear of the teeth (TFW) for various loading and lubrication conditions. The overall classification efficiency of ID3 algorithm and rough set method are compared and the results are presented.

2. Experimental Studies

The fault simulator with sensor is shown in Figure 1 and details of the bevel gear box are shown in Figure 2. A variable speed DC motor (0.5 hp) with a maximum speed of 3000 rpm is the basic drive. A short shaft of 30 mm diameter is attached to the shaft of the motor through a flexible coupling; this is to minimize effects of misalignment and transmission of vibration from the motor. The shaft is supported at its ends by two roller bearings. From this shaft the motion is transmitted to the bevel gear box by means of a belt drive. The gear box is of dimensions 150 mm X 170 mm X 120 mm, the full lubrication level is 110 mm, and half lubrication level is 60 mm. SAE 40 oil was used as the lubricant. An electromagnetic spring loaded disc brake was used to load the gear wheel. A torque of 8 N m was applied at the full load condition. The various defects are created in the pinion wheels and the mating gear wheel is not disturbed. With the sensor mounted on top of the gear box, vibration signals are obtained for various conditions. The selected area is made flat and smooth to ensure effective coupling. A piezoelectric accelerometer (Dytran model) is mounted on the flat surface using a direct adhesive mounting technique. The accelerometer is connected to the signal-conditioning unit (DACTRAN FFT analyzer), where the signal goes through the charge amplifier and an Analogue-to-Digital Converter (ADC). The vibration signal in digital form is fed to a computer through a USB port. The software, RT Pro-series, that accompanies the signal conditioning unit is used for recording the signals directly in the computer’s secondary memory. The signal is then read from the memory and replayed and processed to extract relevant features.

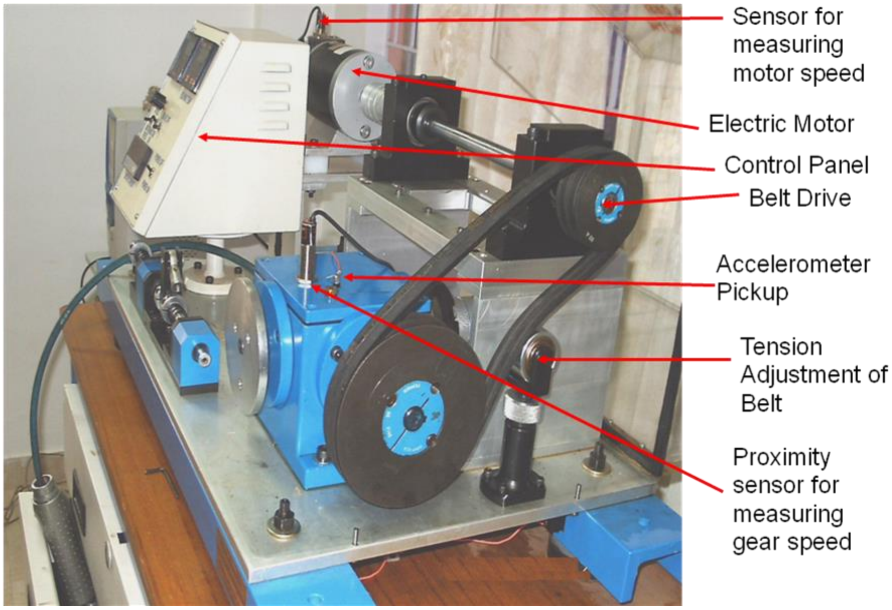

Figure 1. Fault Simulator Setup

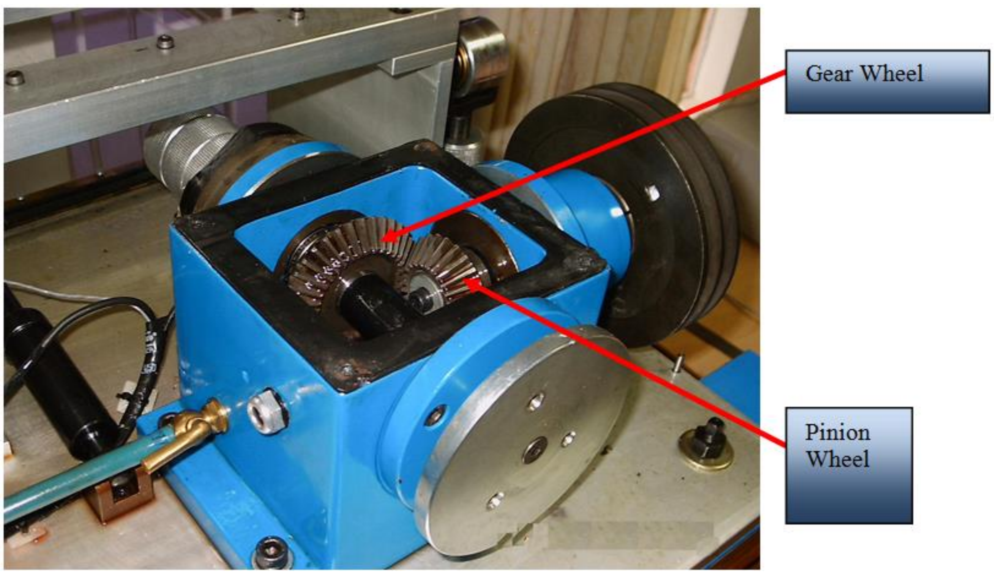

Figure 2. Inner View of Bevel Gear Box

2.1. Experimental Procedure

In the present study, four pinion wheels whose details are as mentioned in Table 1 were used. One was a new wheel and was assumed to be free from defects. In the other three pinion wheels, defects were created using Electric Discharge Machining in order to keep the size of the defect under control. The details of the various defects are depicted in Table 2 and their views are shown in Figure 3(a) to Figure 3(c). The size of the defects is a little larger than one would normally encounter in practical situation; however, it is in line with work reported in the literature [27]. The vibration signal from the piezoelectric pickup mounted on the test bearing was taken, after allowing initial running of the gear box for some time.

Table 1. Details of Faults under Investigation

Gears	Fault Description	Dimension (mm)
G1	Good	--
G2	Gear Tooth Breakage (GTB)	8
G3	Gear with Crack at root (GTC)	0.8 x 0.5 x 20
G4	Gear with Face Wear (TFW)	0.5

The sampling frequency was 12,000 Hz and sample length was 8192 for all speeds and all conditions. The sample length was essentially chosen arbitrarily. However, the following points were considered, as statistical measures are more meaningful with larger sample sizes.

Table 2. Gear Wheel and Pinion Details

Parameters	Gear Wheel	Pinion Wheel
No. of teeth	35	25
Module	2.5	2.5
Normal pressure angle	20^0	20^0
Shaft angle	90^0	90^0
Top clearance	0.5 mm	0.5 mm
Addendum	2.5 mm	2.5 mm
Whole depth	5.5 mm	5.5 mm
Chordal tooth thickness	$3.93^{-0.150}$ mm	$3.92^{-0.110}$ mm
Chordal tooth height	2.53 mm	2.55 mm
Material	EN8	EN8

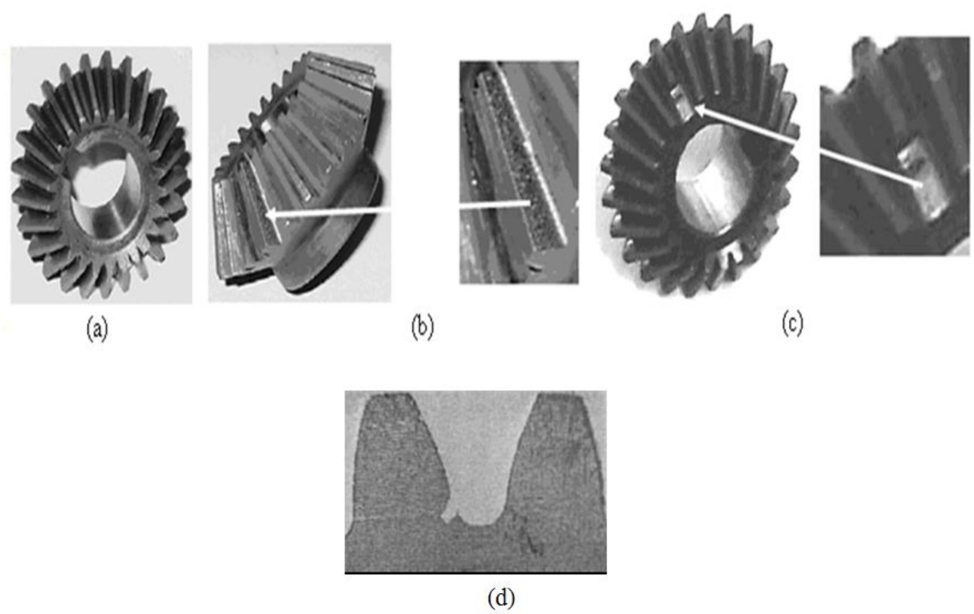

Figure 3. (a) View of Good Pinion Wheel, (b) Pinion Wheel with Face Wear (TFW), (c) Pinion Wheel with Tooth Breakage (GTB) and (d) Pinion Wheel with Tooth Crack(GTC)

On the other hand, as the number of samples increases so does the computational time. To strike a balance, a sample length of around 10000 was chosen. In some feature extraction techniques, which will be used with the same data, the number of samples is required to be 2^n.

Table 3. Abbreviations and their Description

Abbreviation	Gears Considered	Lubrication	Load
d_noload	Good, GTB, TCW, TFW	Dry	No Load
d_fulload	Good, GTB, TCW, TFW	Dry	Full Load
h_noload	Good, GTB, TCW, TFW	Half	No Load
h_fulload	Good, GTB, TCW, TFW	Half	Full Load
f_noload	Good, GTB, TCW, TFW	Full	No Load
f_fulload	Good, GTB, TCW, TFW	Full	Full Load

The nearest 2^n to 10,000 is 8192 and hence this was chosen as the sample length. Many trials were taken at the set speed to obtain vibration signal data. The raw vibration signals acquired for various experimental conditions from the gear box using FFT are shown in Figure 4(a) to Figure 4(d).

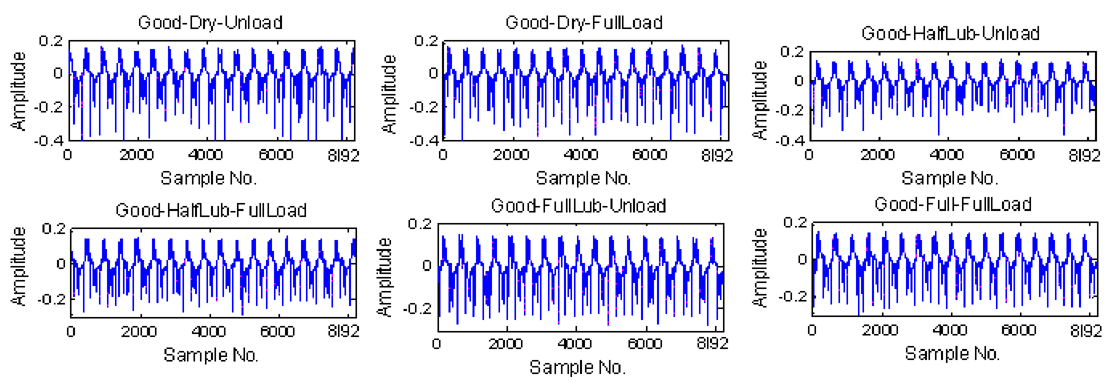

Figure 4(a). Vibration Signals for Good Pinion Wheel (Good) under different Lubrication and Loading Conditions

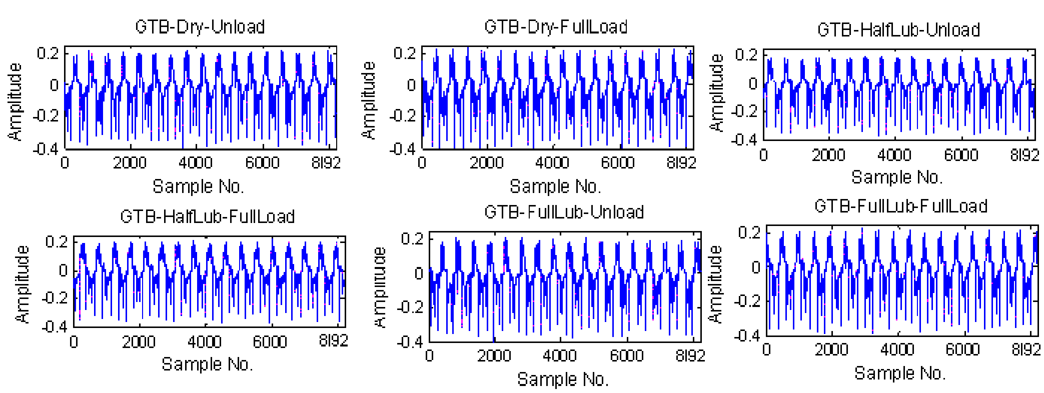

Figure 4(b). Vibration Signals for Pinion Wheel with Teeth Breakage (GTB) under different Lubrication and Loading Conditions

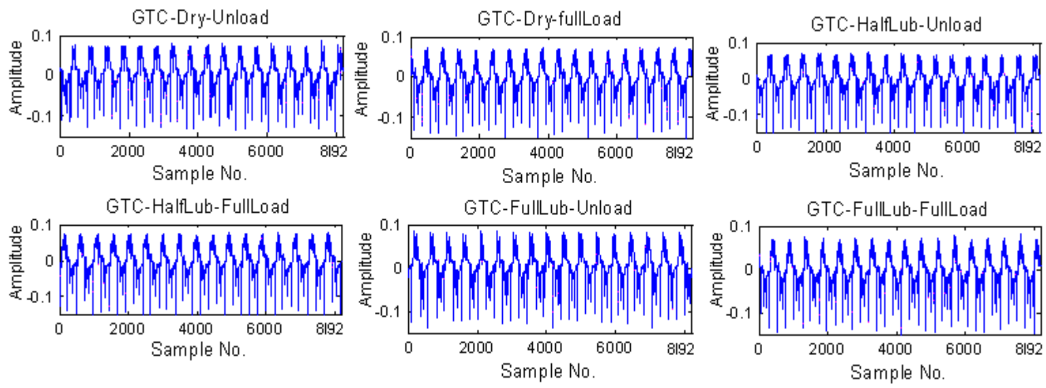

Figure 4(c).Vibration Signals for Pinion Wheel with Crack at Root (GTC) under different Lubrication and Loading Conditions

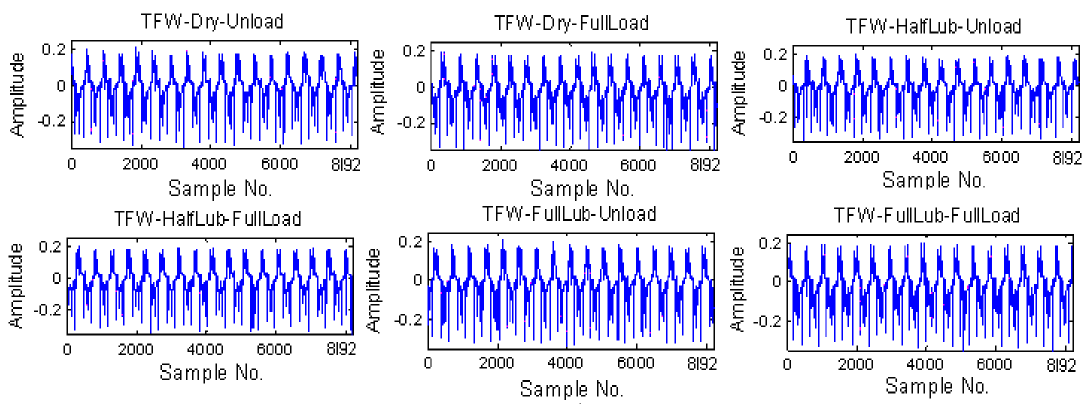

Figure 4(d). Vibration Signals for Pinion Wheel with Teeth Face Wear (TFW) under different Lubrication and Loading Conditions

3. Feature Extraction using Rough Set Method

Rough Sets Theory [28] is a new mathematical tool to handle vagueness and uncertainty inherent in making decisions. Rough Set Theory finds applications primarily in some branches of artificial intelligence and cognitive sciences, such as machine learning, knowledge discovery from databases, expert systems, inductive reasoning, automatic classification, pattern recognition and learning algorithm. Since Z. Pawlak proposed the RST in 1982, this theory has been well studied by many researchers and has made great progress. Due to its advantage, which include the elimination of the need for additional information about data and the ability to extract rules directly from data itself, this theory has been used in more and more domains [29]. In the methodology, a database is regarded as a decision table, which is made up of the universe of discourse, a family of equivalence relations over the universe, condition attributes and decision attributes. The rule discovery process in rough set analysis involves simplifying the decision tables with elimination of superfluous attributes and values of attributes, and determining simple rules relating condition and decision attributes. Since its introduction, the rough set method has increasingly been applied to derive rules, to provide reasoning and to discover relationships in qualitative, incomplete, or imprecise data. This capability is especially important for fault diagnosis. In the following section, the fundamental knowledge about RST is introduced.

3.1. Information Systems

The process of dividing the universe of objects into different categories is called classification; RST deals with the analysis of this classificatory property of a set of objects. Large datasets acquired from measurements or from human experts may, for instance, represent vague, uncertain or incomplete knowledge. RST provides the means to discern and classify objects in datasets of this type, when it is not possible to divide the objects into defined categories.

In RST knowledge is represented in information systems. An information system is a dataset, such as that represented in Table 6. Each row in the table represents an object, for instance a case or an event. Each column in the table represents an attribute, for instance, a variable, an observation or a property. To each object (row) there are assigned some attribute values.

An information system is defined as

$$(= (U, A) \quad (1)$$

U is the nonempty finite set of objects called the Universe.
A the nonempty finite set of attributes such that

$\alpha : U \rightarrow V\alpha$ for every $\alpha \in A$, where $V\alpha$ is the value of set α .

The information system is shown in Table 6. The objects (rows) in the table are statistical features for different fault conditions of the gear box. Different attributes (columns) are measured for each fault condition. The measured attributes are mean, standard error, median, standard deviation, standard variance, kurtosis, skewness, range, minimum, maximum and sum.

3.1.1. Indiscernibility

One of the most important concepts of RST is indiscernibility, which is used to define equivalence classes for the objects. Given a subset of attributes $B(A,$ each such subset defines an equivalence relation *INDA*(*B*) called an indiscernibility relation. This indiscernibility relation is defined as

$$\text{INDA(B)} = \{(x, x') \in U^2 \; ((\alpha \in B, \alpha(x) = \propto (x')\} \quad (2)$$

where *x* and *x*' are objects in *A*.

Eq. (2) states that the subset of attributes, B, will define the partition of the universe into sets such that each object in a set cannot be distinguished from other objects in the set using only the attributes in B. The sets into which the objects are divided are called equivalence classes.

3.1.2. Decision Systems

In the information system the new attribute represents some classification of the objects. The system is called a decision system:

$$(= (U, A \cup \{d\}) \,,$$

where *d* is the decision attribute. The elements of *A* are called conditional attributes or conditions.

The decision is not necessarily constant on the equivalence classes. That is, for two objects belonging to the same equivalence class, the value of the decision attributes may be different. In this case, the decision system is inconsistent (non-deterministic). If a unique classification can be made for all the equivalence classes, the system is consistent (deterministic). The decision attribute is introduced in the information system. This decision attribute states the good and different faulty conditions of the gear box. The decision system is shown in Table 6.

3.1.3. Set Approximation

In order to classify an object based only on the equivalence class to which it belongs, the concept of set approximation is used. Given an information system, $(= (U, A)$, and a subset of attribute, $B (A$, to approximate a set of objects, *X*, using only the information contained in *B*:

$$\underline{B}X = \{x | [x]_B (X\} \tag{3}$$

$\underline{B}\mathbf{X}$ is the lower approximation of *X;*

$$\overline{B}X = \{x | [x]_B (X \neq \mathbf{0}\} \tag{4}$$

where $\overline{B}X$ is the upper approximation of *X*.

The lower approximation is the set containing all objects for which the equivalence class corresponding to the object is a subset of the set. This set contains all objects which certainly belong to the set *X*.

The upper approximation is the set containing the objects for which the intersection of the object's equivalence class and the set to be approximated is not the empty set. This set contains all objects which possibly belong to the set X. Then the boundary region is defined:

$$BN_B(X) = \bar{B}X - \underline{B}X \quad (5)$$

3.1.4. Reducts

In most of the cases not all of the knowledge in an information system is necessary to divide the objects into classes. In these cases, it is possible to reduce knowledge. Reducing the knowledge results in *reducts*. A reduct is a minimal set of attributes. $B (A$, such that

$$IND_A(B) = IND_A(A) \quad (6)$$

A total of 11 reducts were obtained from the decision table and are given in Table 4.

3.1.5. Discernibility Matrices

A discernibility matrix of (is a symmetric $n \times n$ matrix with entries

$C_{ij} = \alpha \in A | \alpha(X_i) \neq \propto (X_j)$ for i, j = 1 to n.

The entries for each object are thus the attributes that are needed in order to discern object i from object j.

3.1.6. Discernibility Functions

A discernibility function can be built from the discernibility matrix. A discernibility function f_α for an information system (is a Boolean function of *m* Boolean variables $\propto_1^*, \ldots, \propto_n^*$ defined as below, where $c_{ij}^* = \{\propto^* \ | \alpha \in c_{ij}\}$,

$$f_\downarrow A \ (\alpha_\downarrow 1^\uparrow *, \ldots, \alpha_\downarrow M^\uparrow *) \ = \ (\{"(" \ c_\downarrow IJ^\uparrow * \ \vdash| 1 \leq j \leq i \dashv \leq n, c_\downarrow ij \ \neq 0\} \quad (7)$$

The discernibility function is a conjuction of all the entries in the discernibility matrix that are not the empty set. The conjuction may, if possible, be simplified. The results of simplification are the possible reducts for the information system. It is also possible to generate a discernibility function from the discernibility matrix for one of the objects in the information system.

Table 4. Condition Attributes and their Frequency in Reducts

(a)

Condition	**Dry_No Load**		**Dry_Full Load**	
Attribute	**Frequency**	**Frequency (%)**	**Frequency**	**Frequency (%)**
Standard Error	7.1	7.1	1	5.3
Mean	14.3	14.3	3	15.8
Median	14.3	14.3	3	15.8
Range	14.3	14.3	1	5.3
Standard Deviation	7.1	7.1	1	5.3
Variance	7.1	7.1	1	5.3
Kurtosis	7.1	7.1	1	5.3
Skewness	7.1	7.1	3	15.8
Minimum	7.1	7.1	3	15.8
Maximum	7.1	7.1	1	5.3
Sum	7.1	7.1	1	5.3

This is done by considering only one row (or column) in the discernibility matrix, and forming a conjuction of all the entries in the row (or column). The possible reducts for the particular object are obtained by simplifying this conjunction.

The role of different attributes in reducts for different conditions of the gear are given in Table 4a to 4c. Also highlighted in these tables are the attributes which contribute more in the reducts for the fault identification and these reducts are used further to classify the faults after testing the rules. Frequency and frequency percentage is the number of times that a particular attribute contributes to the reducts. It is also shown in the table. Frequency is the measure of the "strength" of the reduct in studying and classifying the different conditions of the gear under investigation.

(b)

Condition	**Half Lub_No Load**		**Half Lub_Full Load**	
Attribute	**Frequency**	**Frequency (%)**	**Frequency**	**Frequency (%)**
Standar d Error	1	2.9	1	5.9
Mean	5	14.3	2	11.8
Median	5	14.3	3	17.6
Skewness	5	14.3	1	5.9

Range	5	14.3	3	17.6
Minimum	5	14.3	1	5.9
Maximum	5	14.3	1	5.9
Standard Deviation	1	2.9	1	5.9
Variance	1	2.9	1	5.9
Kurtosis	1	2.9	2	11.8
Sum	1	2.9	1	5.9

(c)

Condition	**Full Lub_No Load**		**Full Lub_Full Load**	
Attribute	**Frequency**	**Frequency (%)**	**Frequency**	**Frequency (%)**
Standard Error	1	4.2	1	7.1
Mean	3	12.5	2	14.3
Kurtosis	4	16.7	1	7.1
Skewness	4	16.7	1	7.1
Maximum	4	16.7	1	7.1
Standard Deviation	1	4.2	1	7.1
Median	3	12.5	2	14.3
Variance	1	4.2	1	7.1
Range	1	4.2	1	7.1
Minimum	1	4.2	2	14.3
Sum	1	4.2	1	7.1

Table 5. Reducts

Reducts					
Dry_Noload	Dry_Full Load	Half Lub_No load	Half Lub_full Load	Full Lub_No Load	Full Load_Full Load
Mean	Mean	Mean	Mean	Mean	Mean
Median	Median	Median	Median	Median	Median
Range	Skewness	Range	Range	Skewness	Minimum
	Minimum	Skewness	Kurtosis	Maximum	

Table 5 gives the details about the reducts selected for different conditions of the gear box which play a dominant role in classifying the various faults of gear the under different conditions of study. It can be seen from the table that mean, median, range, minimum, range, skewness and kurtosis are the salient statistical features that contribute in diagnosing the various faults of the gear.

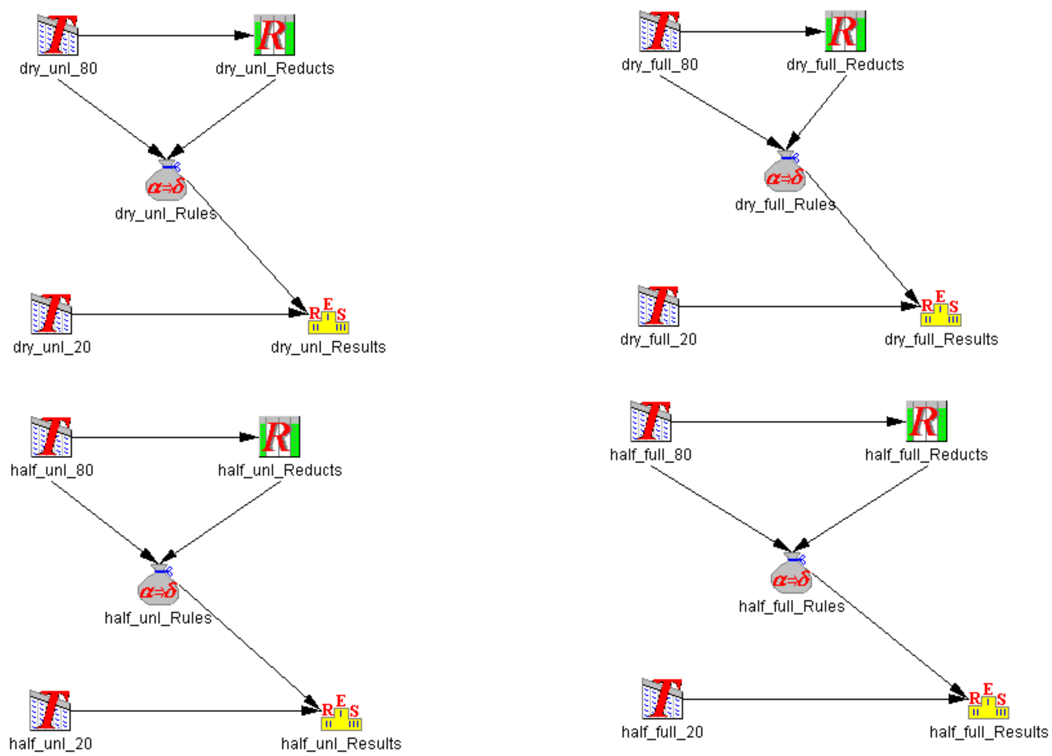

Figure 5. Flow Diagram of Classification using Reducts

Figure 5 illustrates how classifications are carried out for various conditions of the gear. Based on the training, the rules are formed and tested with test data, and the final decisions are stored. 80 out of 100 datasets were used for training and obtaining the reducts and the remaining 20 datasets were tested and the results stored.

Table 6. Sample Information and Decision System

Mean	Standard Error	Median	Standard Deviation	Sample Variance	Kutrosis	Skewness	Range	Minimum	Maximum	Sum	Condition
-2.17E-05	0.000827	-0.002	0.074828	0.005599	1.548892	-0.4598	0.600599	-0.42594	0.174663	-0.17769	Good_Dry_NoLoad
-3.46E-05	0.001132	-0.00537	0.102489	0.010504	0.064869	-0.13106	0.663076	-0.43644	0.226631	-0.28328	GTB_Dry_NoLoad
0.000272004	0.000387	-0.00127	0.035036	0.001228	-0.02859	-0.1433	0.221853	-0.14293	0.078923	2.228254	TCW_Dry_NoLoad
-0.00056132	0.000953	-0.00267	0.086265	0.007442	0.435255	-0.21089	0.648956	-0.44327	0.205686	-4.59835	TFW_Dry_NoLoad
1.42E-05	0.000803	-0.00209	0.072648	0.005278	0.8822	-0.30272	0.524239	-0.35391	0.170326	0.116516	Good_Dry_FullLoad
-0.00029674	0.001272	-0.00455	0.115103	0.013249	-0.3196	-0.06682	0.82621	-0.56434	0.261866	-2.43094	GTB_Dry_FullLoad
-2.79E-05	0.000399	-0.00176	0.036074	0.001301	-0.11145	-0.12784	0.233282	-0.15545	0.077828	-0.22895	TCW_Dry_FullLoad
-0.00015297	0.001004	-0.00284	0.090908	0.008264	-0.09175	-0.09685	0.624796	-0.42334	0.201451	-1.25316	TFW_Dry_FullLoad
0.000387509	0.00072	-0.00084	0.065148	0.004244	0.506634	-0.19964	0.505958	-0.34431	0.161646	3.174477	Good_Half Lub_No Load
0.000333304	0.001019	-0.00448	0.092199	0.008501	-0.0072	-0.13242	0.563432	-0.36728	0.196155	2.730427	GTB_Half Lub_No Load
-7.79E-05	0.0004	-0.00256	0.036232	0.001313	0.216011	-0.2009	0.246506	-0.17077	0.075736	-0.63778	TCW_Half Lub_No Load
-0.00030281	0.00093	-0.00253	0.08413	0.007078	0.311413	-0.14833	0.541595	-0.33657	0.205028	-2.48062	TFW_Half Lub_No Load
-0.00025451	0.000748	-0.00138	0.067736	0.004588	0.03183	-0.12238	0.49762	-0.34622	0.1514	-2.08495	Good_Half Lub_Full Load
0.00064088	0.001119	-0.00239	0.10124	0.01025	-0.43965	-0.05562	0.590631	-0.36841	0.222218	5.250087	GTB_Half Lub_Full Load
-7.22E-05	0.000425	-0.00219	0.038454	0.001479	-0.06152	-0.13875	0.25718	-0.1739	0.083276	-0.59136	TCW_Half Lub_Full Load
0.000350947	0.000987	-0.00054	0.089297	0.007974	-0.05702	-0.08254	0.562373	-0.36079	0.201585	2.874958	TFW_Half Lub_Full Load
-0.00015069	0.000723	-0.00306	0.065448	0.004283	0.272113	-0.15057	0.424623	-0.26985	0.154772	-1.23448	Good_Full Lub_No Load
-0.00030237	0.00108	-0.00555	0.097765	0.009558	0.007557	-0.10409	0.670243	-0.43203	0.238217	-2.47708	GTB_Full Lub_No Load
-0.00013892	0.000373	-0.00182	0.033768	0.00114	0.440978	-0.1801	0.2288	-0.14537	0.083426	-1.13806	TCW_Full Lub_No Load
0.001001465	0.000944	-0.00263	0.085464	0.007304	0.365172	-0.15333	0.573021	-0.36521	0.207816	8.203998	TFW_Full Lub_No Load
-1.85E-05	0.000758	-0.00221	0.068576	0.004703	0.058083	-0.09433	0.454113	-0.29366	0.160453	-0.15174	Good_Full Lub_Full Load
-0.00049254	0.001224	-0.00407	0.110791	0.012275	-0.38472	-0.02616	0.82311	-0.54687	0.27624	-4.03488	GTB_Full Lub_Full Load
0.000268614	0.000399	-0.00073	0.036157	0.001307	0.011868	-0.11223	0.234023	-0.15397	0.08005	2.20049	TCW_Full Lub_Full Load
-0.00070173	0.001004	-0.0016	0.090852	0.008254	0.056172	-0.09484	0.63638	-0.42378	0.212602	-5.7486	TFW_Full Lub_Full Load

Table 7. Classification of Gear Faults using Reducts

(a)

Condition	Dry No Load							
Actual	Predicted							
		Good	GTB	TCW	TFW	No. of Objects	Accuracy	Coverage
	Good_Dry_No Load	20	0	0	0	20	1	1
	GTB_Dry_No Load	0	8	0	0	20	1	0.4
	TCW_Dry_No Load	0	0	20	0	20	1	1
	TFW_Dry_No Load	0	0	0	20	20	1	1
Total Accuracy	1							
Total Coverage	0.85							

(b)

Condition	Dry_Full Load							
Actual	Predicted							
		Good	GTB	TCW	TFW	No. of Objects	Accuracy	Coverage
	Good_Dry_Full Load	19	0	0	0	20	1	0.95
	GTB_Dry_Full Load	0	11	0	0	20	1	0.55
	TCW_Dry_Full Load	0	0	3	0	20	1	0.15
	TFW_Dry_Full Load	0	0	0	13	20	1	0.65
Total Accuracy	1							
Total Coverage	0.575							

(c)

Condition	Half Lub_No Load							
Actual	Predicted							
		Good	GTB	TCW	TFW	No. of Objects	Accuracy	Coverage
	Good_Half Lub_No Load	20	0	0	0	20	1	1
	GTB_Half Lub_No Load	0	19	0	0	20	1	0.95
	TCW_Half Lub_No Load	0	0	20	0	20	1	1
	TFW_Half Lub_No Load	0	0	0	19	20	1	0.95
Total Accuracy	1							
Total Coverage	0.975							

(d)

Condition	Half Lub_Full Load							
Actual	Predicted							
		Good	GTB	TCW	TFW	No.of Objects	Accuracy	Coverage
	Good_Half Lub_Full Load	18	0	0	0	20	1	0.9
	GTB_Half Lub_Full Load	0	10	0	0	20	0.909	0.55
	TCW_Half Lub_Full Load	0	0	20	1	20	1	1
	TFW_Half Lub_Full Load	0	0	0	6	20	1	0.33
Total Accuracy	0.982							
Total Coverage	0.688							

(e)

Condition	Full Lub_No Load							
Actual	Predicted							
		Good	GTB	TCW	TFW	No. of Objects	Accuracy	Coverage
	Good_Full Lub_No Load	19	0	0	0	20	1	0.95
	GTB_Full Lub_No Load	0	17	0	0	20	1	0.85
	TCW_Full Lub_No Load	0	0	20	0	20	1	1
	TFW_Full Lub_No Load	0	0	0	20	20	1	1
Total Accuracy	1							
Total Coverage	0.85							

(f)

Condition	Full Lub_Full Load							
Actual	Predicted							
		Good	GTB	TCW	TFW	No. of Objects	Accuracy	Coverage
	Good_Full Lub_Full Load	15	0	0	0	20	1	0.75
	GTB_Full Lub_Full Load	0	18	0	0	20	1	0.9
	TCW_Full Lub_Full Load	0	0	20	0	20	1	1
	TFW_Full Lub_Full Load	0	0	0	8	20	1	0.4
Total Accuracy	1							
Total Coverage	0.762							

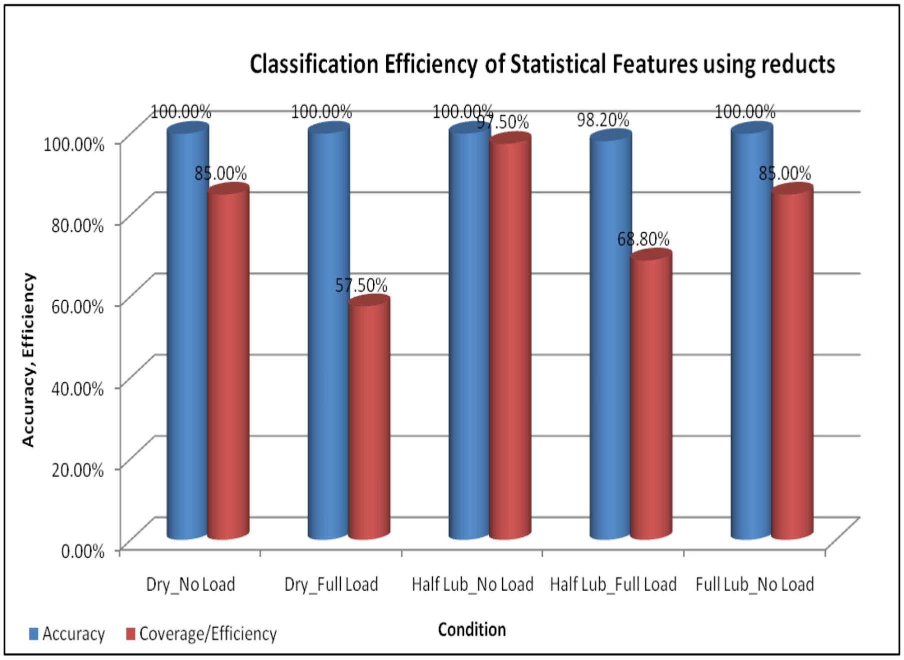

Figure 6. Overall Classification Efficiency of Gear Faults using Reducts

Table 7(a) to 7(f) shows the results of classification of different gear faults using reducts. The table presenting the classification results provides a variety of information. The central part is occupied by the confusion matrix. Rows in this matrix correspond to actual decision classes (all possible values of decision) while columns represent decision values as returned by the classifier while classifying the data. Table 8(a) shows that the reducts predict the different gear faults in dry-no-load condition accurately, except for gear tooth breakage (GTB). Only 8 instances are predicted accurately out of a total of 20 objects. Total accuracy is the ratio of the number of correctly classified cases to the number of all tested cases. The total accuracy is 1 for this case. Total coverage that were recognised by the classifier in the case of gear tooth breakage is around 0.4. However the overall coverage accuracy for the dry-no-load condition of the gear box is as high as 0.85. The other tables give the results for other conditions of the gear box. Figure 6 gives the overall classification efficiency of the reducts selected. This shows that, overall RST has a great capability in studying and predicting different faults of the gear under investigation. For all the conditions of the gear box the classification efficiency is 100%, but the coverage efficiency is a maximum of 98.2% in the case of half lubrication, full load condition. In the case of dry lubrication full load condition, the coverage efficiency is as low as 57.50%. This may be attributed to the smaller number of instances or the strength of the reducts are not good. But overall the performace of the reducts, and thereby RST seems to be promising for fault diagnosis of the gear box under investigation.

4. Using ID3 Algorithm for Gear Fault Classification

ID3 is a tree-based knowledge representation methodology used to represent classification rules. A standard tree induced with c5.0 (or possibly ID3 or c4.5) consists of a number of branches, one root, a number of nodes and a number of leaves. One branch is a chain of nodes from the root to a leaf; and each node involves one attribute. The occurrence of an attribute in a tree provides the information about the importance of the associated attribute [29]. The procedure of forming the Decision Tree and exploiting the same for feature selection is characterized by the following.

1. The available set of features forms the input to the algorithm; the output is the Decision Tree.

2. The Decision Tree has leaf nodes, which represent class labels, and other nodes associated with the classes being classified.

3. The branches of the tree represent each possible value of the feature node from which they originate.

4. The ID3 can be used to classify feature vectors by starting at the root of the tree and moving through it until a leaf node, which provides a classification of the instance, is identified.

5. At each decision node in the Decision Tree, one can select the most useful feature for classification using appropriate estimation criteria. The criterion used to identify the best feature invokes the concepts of entropy reduction and information gain – discussed in the following sub section.

4.1. Information Gain and Entropy Reduction

Information gain measures how well a given attribute separates the training examples according to their target classification. The measure is used to select among the candidate features at each step while growing the tree. Information gain is the expected reduction in entropy caused by portioning the samples according to this feature.

Information gain (*S*, *A*) of a feature *A* relative to a collection of examples *S*, is defined as:

$$Gain(S,A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

[30]

where, *Values (A)* is the set of all possible values for attribute *A*,

S_v is the subset of *S* for which feature *A* has value *v* (i.e., $S_v = \{s \in S \mid A(s) = v\}$).

Note the first term in the equation for *Gain* is just the entropy of the original collection *S* and the second term is the expected value of the entropy after S is partitioned using feature *A*. The expected entropy described by the second term is simply the sum of the entropies of each subset S_v, weighted by the fraction of samples $|S_v|/|S|$ that belong to S_v. *Gain (S,A)* is therefore the expected reduction in entropy caused by knowing the value of feature *A*. Entropy is a measure of homogeneity of the set of examples and is given by

$$Entropy(S) = \sum_{i=1}^{c} - P_i \log_2 P_i \tag{8}$$

The algorithm identifies the good features for the purpose of classification from the given training data set, and thus reduces the domain knowledge required to select good features for the pattern classification problem. The decision trees shown in Figure 7a to Figure 7f are for various lubrication and loading conditions of different gear faults compared with good conditions of the pinion gear wheel.

Based on the trees it is clear that of all the statistical features, standard error plays a dominant role in fault classification. The decision tree algorithm with the help of a single statistical parameter is able to diagnose the various gear faults better than that requiring 11 statistical parameters required from RST. These features are used for training and testing the algorithm and the test results are tabulated in Table 8a to 8f.

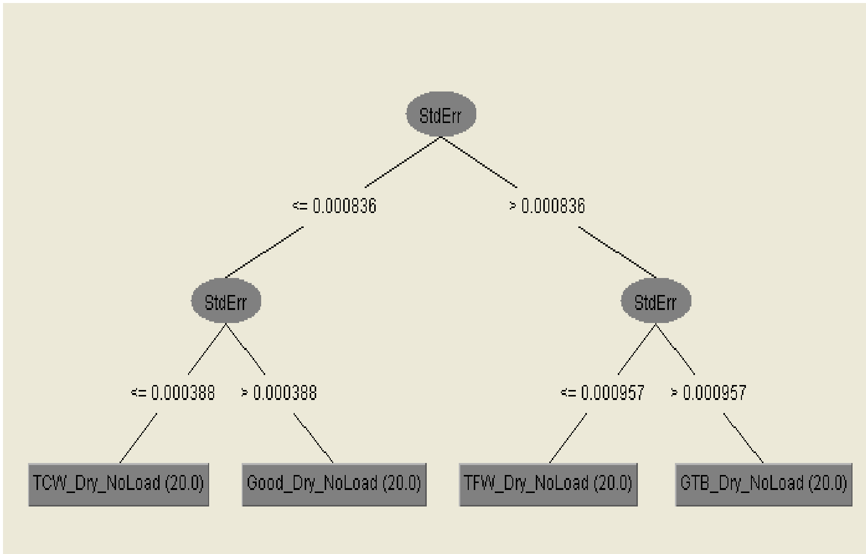

Figure 7a. Good-Dry-No Load Vs GTB, GTC, TFW-Dry-No Load

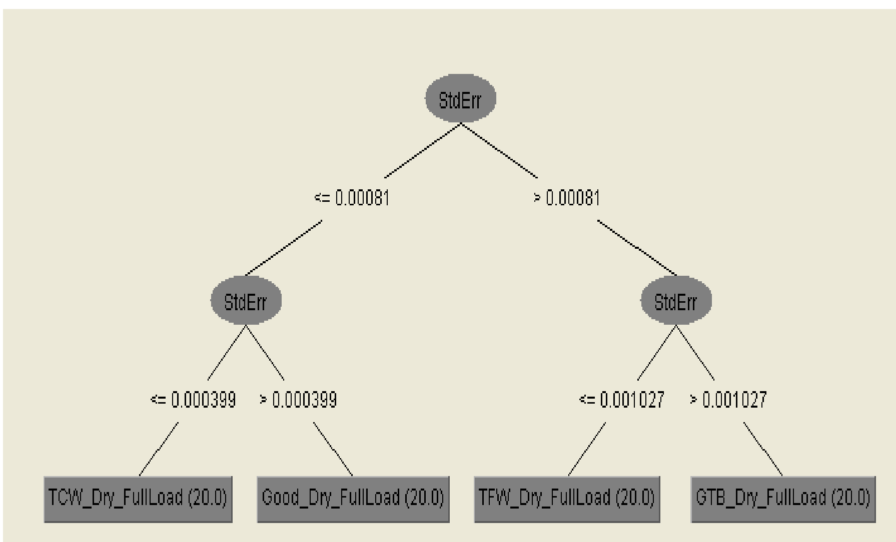

Figure 7b. Good-Dry-Full Load Vs GTB, GTC, TFW-Dry-Full Load

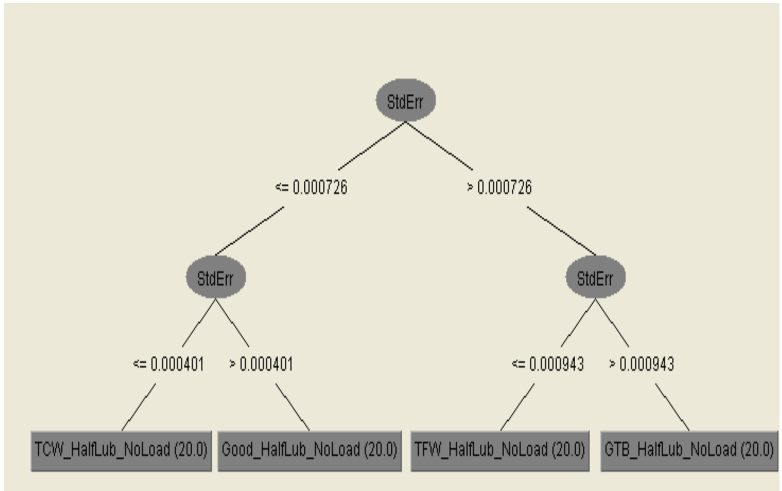

Figure 7c. Good-Half-Lub-No Load Vs GTB, GTC, TFW-Half-Lub-No Load

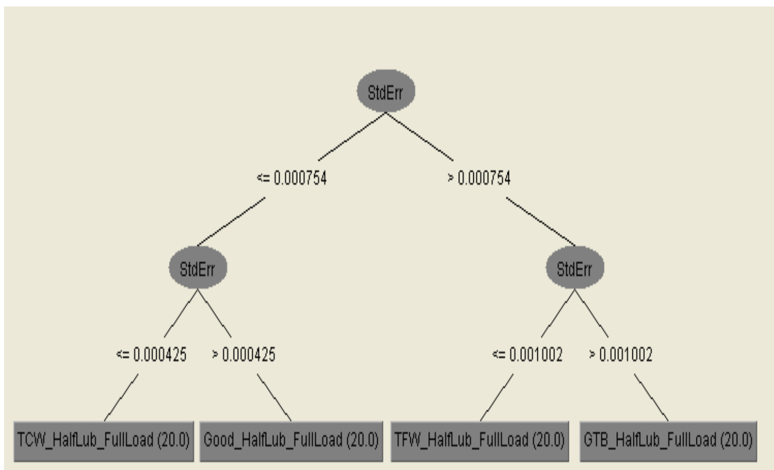

Figure 7d. Good- Half-Lub -Full Load Vs GTB, GTC, TFW- Half-Lub -Full Load

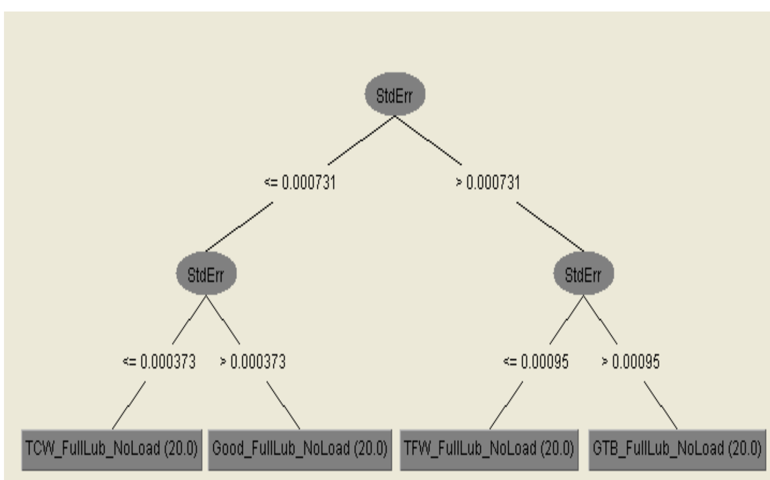

Figure 7e. Good-Full-Lub-No Load Vs GTB, GTC, TFW-Full-Lub-No Load

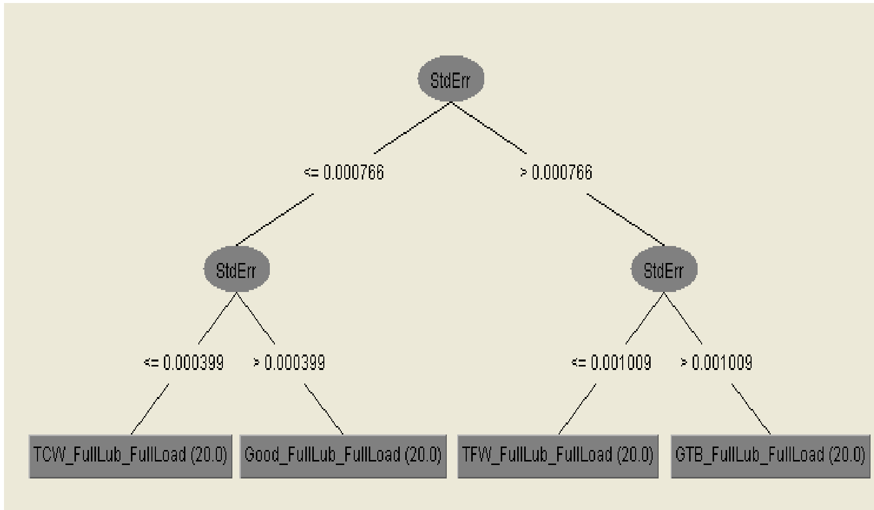

Figure 7f. Good- Full-Lub -Full Load Vs GTB, GTC, TFW- Full-Lub -Full Load

Table 8. Classification of Gear Faults using Decision Tree

(a)

Condition	Dry No Load				
Actual	**Predicted**				
		Good	**GTB**	**TCW**	**TFW**
	Good_Dry_No Load	99	0	0	1
	GTB_Dry_Nol Load	0	100	0	0
	TCW_Dry_No Load	0	0	100	0
	TFW_Dry_No Load	0	0	0	100

(b)

Condition	Dry Full Load				
Actual	**Predicted**				
		Good	**GTB**	**TCW**	**TFW**
	Good_Dry_Full Load	100	0	0	0
	GTB_Dry_Full Load	0	100	0	0
	TCW_Dry_Full Load	1	0	99	0
	TFW_Dry_Full Load	0	0	0	100

(c)

Condition	Half Lubrication No Load				
Actual	**Predicted**				
		Good	**GTB**	**TCW**	**TFW**
	Good_Half Lub_No Load	99	0	0	1

	GTB_Half Lub_Nol Load	0	100	0	0
	TCW_Half Lub_No Load	0	0	100	0
	TFW_Half Lub_No Load	0	1	0	99

(d)

Condition	Half Lubrication Full Load				
Actual	**Predicted**				
		Good	**GTB**	**TCW**	**TFW**
	Good_Half Lub_Full Load	99	0	0	1
	GTB_Half Lub_Full Load	0	100	0	0
	TCW_Half Lub_Full Load	0	0	100	0
	TFW_Half Lub_Full Load	0	1	0	99

(e)

Condition	Full Lubrication No Load				
Actual	**Predicted**				
		Good	**GTB**	**TCW**	**TFW**
	Good_Full Lub_No Load	99	0	0	1
	GTB_Full Lub_Nol Load	0	100	0	0
	TCW_Full Lub_No Load	0	0	100	0
	TFW_Full Lub_No Load	0	0	0	100

(f)

Condition	Full Lubrication Full Load				
Actual	**Predicted**				
		Good	**GTB**	**TCW**	**TFW**
	Good_Full Lub_Full Load	99	0	0	1
	GTB_Full Lub_Full Load	0	100	0	0
	TCW_Full Lub_Full Load	1	0	99	0
	TFW_Full Lub_Full Load	0	1	0	99

Table 8a to 8f shows the results of classification of different gear faults using ID3 reducts. The table presenting the classification results provides a range of information. The central part is occupied by the confusion matrix. Rows in this matrix correspond to actual decision classes (all possible values of decision) while column represent decision values as returned by the classifier while classification. Table 8a shows that the algorithm predicts the different gear faults in dry-no-load condition accurately except gear tooth face wear (TFW). But only one instance of good condition is predicted as tooth face wear. Other faults are classified or diagnoised accurately. In case of Table 8b, it can be seen that all the falts of the gear are classified accurately except one instance of misclassification. 1 instance of tooth crack is misclassified as gear tooth wear. The same interpretation can be made for the other tables. Based on all the above tables it is found that the ID3 algorithm classifies all the faults of the gear under investigation with a high level of accuracy.

5. Discussion

The use of RST and ID3 algorithm for statistical features extracted from the vibration signatures of the gear box was found to be very efficient for classification of the faults of the gear box. So far the traditional methods were employed to detect the one fault category of gears. But here in this work RST and ID3 algorithm is employed for classifying, thereby diagnosing more than one faults of the gears under investigation.

Out of all statistical features it was found that mean, median, range, minimum, range, skewness and kurtosis are the salient statistical features that contribute in fault diagnosing the various faults of the gear using RST. In the case of the ID3 algorithm, standard error plays a dominant role in classifying the various faults. It helps to reduce the number of data to be processed for the condition monitoring of the gear box, thereby reducing the processing time to a great extent. The key success in condition monitoring any equipment lies in the ease and fastness of arriving at the decisions. This work reports that both RST and ID3 algorithm have a promising potential in arriving at decisions on the conditions of the component under investigation. Comparing both these methods, the ID3 algorithm is superior to the RST method in the amount of data to be handled and the time required at arriving at the decisions, which is very crucial in condition monitoring.

6. Conclusion

The present paper dealt with the application of the rough sets idea and ID3 algorithm to fault classification of gear boxes. A measure of classification accuracy, which can serve also as a criterion for feature selection, was developed in order to be used in condition monitoring of the component chosen for study. Feature selection via RST and the D3 algorithm was illustrated using real data from the vibration signatures of the component considered for study. Fault diagnosis of gear box is a core research area in the field of condition monitoring of rotating machines. A comparative study of the classifying ability of RST method and the ID3 algorithm was performed. It was found that ID3 algorithm performs significantly better than the RST method, in the area of quickness at arriving the decisions and data to be processed, which is very crucial in condition monitoring.

References

[1] P. Gadd and P. J. Mitchell, "Condition monitoring of helicopter gearboxes using automatic vibration analysis techniques", AGARD CP 369, Gears and power transmission system for helicopter turboprops, 1984, pp. 29/1-29/10.
[2] J. F. A. Leblanc, J. R. F. Dube and B. Devereux, "Helicopter gearbox vibration analysis in the Canadian Forces—applications and lessons", Proceedings of the First International Conference, Gearbox noise and vibration, IMechE, Cambridge, UK, C404/023, **(1990)**, pp. 173-177.
[3] B. G. Cameron and M. J. Stuckey, "A review of transmission vibration monitoring at Westland Helicopter Ltd.", Proceedings of the 20th European Rotorcraft Forum, vol. 116, **(1994)**, pp. 16/1-116/16.
[4] D. Huang, "Vibration analysis and fault diagnosis for diesel engine", Journal of Vibration Engineering, vol. 8, no. 2, **(1995)**, pp. 144-149.

[5] S. Tsumoto and H. Tanaka, "Automated discovery of medical expert system rules from clinical databases based on Rough Sets", Proceedings of the Second International Conference on Knowledge Discovery and DataMining (KDD'96)., AAAI Press, Menlo Park, **(1996)**, pp. 63-69.
[6] W. Ziarko, R. Golan and D. Edwards, "An application of Data- logic/R knowledge discovery tool to identify strong predictive rules in stock market data", Proceedings of AAAI Workshop on Knowledge Discovery in Databases, (Washington D.C.), **(1993)**, pp. 93-101.
[7] R. Nowicki, R. Slowinski and J. Stefanowski, "Evaluation of vibro acoustic diagnostic symptoms by means of the Rough in Data Mining", Sets Theory, Computers in Industry 20 **(1992)**, pp. 141-152.
[8] R. Slowinski and C. Zopounidis, "Application of the rough set approach to evaluation of bankruptcy risk", Intelligent Systems in Accounting, Finance and Management 4, **(1995)**, pp. 27-41.
[9] R. Slowinski, "Intelligent Decision Support", Handbook of Applications and Advances of Rough Sets Theory, Kluwer Academic Publishing, **(1992)**.
[10] M. R. Rebolledo, "Integrating rough sets and situation-based qualitative models for processes monitoring considering vagueness and uncertainty", Engineering Applications of Artificial Intelligenc, vol. 18, no. 5, **(2005)**, pp. 617-632.
[11] H. W. Qing and R. L. Jing, "A rough set-based fault ranking prototype system for fault diagnosis", Engineering Applications of Artificial Intelligence, vol. 17, no. 8, **(2004)**, pp. 909-917.
[12] I. U. Sikder and A. Gangopadhyay, "Managing uncertainty in location services using rough set and evidence theory", Expert Systems with Applications, vol. 32, no. 2, **(2007)**, pp. 386-396.
[13] H. C. Chou, C. H. Cheng and J. R. Chang, "Extracting drug utilization knowledge using self-organizing map and rough set theory", Expert Systems with Applications, vol. 33, no. 2, **(2007)**, pp. 499-508.
[14] Y. W. Xiang, Y. Jie and L. T. Xiao, "Feature selection based on rough sets and particle swarm optimization", Pattern Recognition Letters, vol. 28, no. 4, **(2007)**, pp. 459-471.
[15] A. E. Hassanien, "Fuzzy rough sets hybrid scheme for breast cancer detection", Image and Vision Computing, vol. 25, no. 2, **(2007)**, pp. 172-183.
[16] R. Agrawal and T. Imielinski, "A. Swami, Databasemining: A performance perspective", IEEE Transactions on Knowldge and Data Engineering, vol. 5, no. 6, **(1993)**, pp. 914-925.
[17] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data", Communications of the ACM, vol. 39, no. 11, **(1996)**, pp. 27-34.
[18] U. Fayyad, G. Piatetsky-Shapiro and R. Uthurusamy, "Advances in knowledge discovery and data mining", MIT Press, Cambridge, MA, **(1996)**.
[19] T. Munakata (Ed.), "Knowledge discovery", Communications of the ACM, vol. 42, no. 11, **(1999)**, pp. 26-29.
[20] P. Langley, H. A. Simon, "Applications of machine learning and rule induction", Communications of the ACM, vol. 38, no. 11, **(1995)**, pp. 54-64.
[21] T. Mitchell, "Machine learning and data mining", Communications of the ACM, vol. 42, no. 11, **(1999)**, pp. 30-36.
[22] L. M. Fu, "Knowledge discovery based on neural networks", Communications of the ACM, vol. 42, no. 11, **(1999)**, pp. 47-50.
[23] V. Dhar and A. Tuzhilin, "Abstract-driven pattern discovery in databases", IEEE Transactions on Knowledge and Data Engineering, vol. 5, no. 6, **(1993)**, pp. 926-937.
[24] Y. Pao, "Adaptive Pattern Recognition and Neural Networks", Addison-Wesley, Reading, **(1989)**.
[25] N. Zhong, A. Skowron and S.Ohsuga (Eds.), "Newdirection in rough sets datamining and granular - Soft computing", Proceedings of the Seventh International Workshop, RSFDGrC'99, Yamaguchi, Japan, Springer, **(1999)**, November 9-11.
[26] A. Mrozek, "A new method for discovering rules from examples in expert systems", International Journal of Man Machine Studies, vol. 36, **(1992)**, pp.127-143.
[27] F. A. Andrade, I. and M. N. M. Badi, "A new approach to time-domain vibration condition monitoring: Gear tooth fatigue crack detection and identification by the Kolmogrov-Smirnov test", Journal of Sound and vibration, vol. 240, no. 5, **(2001)**, pp. 909-07.
[28] Z. Pawlak, J. Grzymala-Busse, R. Slowinski and W. Ziarko, "Rough Sets", Communications on the ACM, vol. 38, no. 11, **(1995)**, pp. 89-95.
[29] L. Shen., F. E. H. Tay, L. Qu and Y. Shen, "Fault Diagnosis using Rough Sets Theory", Journal of Computers in Industry, vol. 23, **(2000)**, pp. 61-72.
[30] Y. H. Peng, P. A. Flach, P. Brazdil and C. Soares, "Decision Tree-Based Data Characterization for Meta-Learning", ECML/PKDD Workshop, IDDM, Helsinki, Finland, **(2002)**.

Authors

Saravanan Natarajan, Ph.D, has 14 years of experience of which 01 years are in teaching in engineering colleges in India and Sultanate of Oman. Is an active researcher in field of condition monitoring and reviewer for technical journals. Has 4 years of industrial experience in the field of CAD/CAE. Has carried out many projects for GEA, USA. Has worked in GEA, Louisville, Kentucky state, USA.