Vibration analysis of gearbox fault diagnosis using DWT and statistical features

Setti Suresh*, ***, M Srinivas* and VPS Naidu**

School of Avionics, Institute of Science and Technology, Jawaharlal Nehru Technological University, Kakinada, India.
**Multi Sensor Data Fusion Lab, FMCD, CSIR-National Aerospace Laboratories, Bangalore, India.*
****Corresponding Author: settisuresh1@gmail.com*

ABSTRACT

The gearbox is one of the critical subsystems in any rotating machinery, which plays a significant role in machine-driven power transmission in terms of change in speed and torque. It plays a vital role in patching different industrial functionalities. The advent of developing different gear technologies and the requirement to fulfil the desired mechanical benefits lead to adding more importance to the gearbox health condition monitoring from various types of fault occurrences at an earlier stage. This study presents the vibration analysis of gearbox fault diagnosis using discrete wavelet transform (DWT) and statistical features. It is observed that, using wavelet reconstruction in the fault diagnosis, a better fault classification is achieved. The fault diagnosis has been presented with an emphasis in time-domain followed by two different approaches. The approach-1 is illustrated as windowing of raw signal, feature extraction, and feature classification using support vector machine (SVM). In the approach-2, after windowing the raw signal, each window of original vibration signal has converted into wavelet coefficients reconstructed signals (without leaving the time domain) using discrete wavelet transform (DWT) at different levels of decomposition followed by approach-1. The fault diagnostic accuracy of SVM is presented with 100 Monte-Carlo runs to validate the consistency in the accomplished result. By observing the success rates in two approaches, it is clear that approach-2 with wavelet coefficient's reconstructed signal is providing better classification accuracy, which can be practically deployed to diagnose the gearbox fault.

Keywords: Vibration analysis; Condition monitoring; Fault diagnosis; Data visualization.

INTRODUCTION

The gearbox finds its application with almost every type of industrial machinery. There are different types of gears under usage, but the basic functionality is to easy out the mechanical power transmission with varying speed and torque. The gearbox as a whole contains cascading of gears arranged in a particular fashion to meet the desired speed and torque conversions using diverse gear ratios. Failure of the gearbox leads to the interruption of the ongoing mechanical process, which decreases the efficient production and cost-effectiveness of the industry. As being such a crucial part, the gearbox health-condition should be monitored to get prior information about the initiation of the fault. There are different condition monitoring analyses, that is, thermal, acoustic, and vibration, which are used to

characterize the behavior of the gearbox. Among the three, vibration analysis has been widely adopted as it contains more numerical information.

The vibration analysis of gearbox faults has been diagnosed using the wavelet transform and wavelet coefficients reconstructed signal tool to bring the importance of multiresolution features in the signal. The implementation of wavelet transforms in the analysis of helicopter gearbox distributed and local faults (irrespective of the size of faults) has been outlined in (Wang, W. J., & McFadden, P.

D., 1996). The fault diagnosis of bearing vibration signal based on wavelet packet transform followed by phase space reconstruction and singular value decomposition has been presented in (Fei, S. 2017), and it is stated that, by adding phase space reconstruction in the analysis through wavelet transform followed by SVM classifier, the fault diagnosis ability of bearing vibration signal is improved. The importance of wavelet reconstruction in the rotor fault (broken bars) diagnosis of induction motor has been illustrated in (Cao Zhitong, Chen Hongping, He Guoguang, & Ritchie, E. (n.d.), 2001). The orientation of fault identification and diagnosis of rotating machinery has a wide variety of aspects based on the available data, experimental results, and comparative studies from each individual researcher [Wang, W. J., & McFadden, P. D., 1996, Fei, S. 2017, & Cao Zhitong, Chen Hongping, He Guoguang, & Ritchie, E. (n.d.), 2001]. However, the main aim is to identify and diagnose the faults in the rotating machinery in an effective way.

This paper highlights the vibration analysis of gearbox fault diagnosis using DWT and statistical features. Time-domain statistical features, which are most commonly used in the classification and fault diagnosis of rotating machinery, were estimated from the wavelet coefficient's reconstructed signal and used with SVM classifier to diagnose the gearbox fault between healthy and broken tooth condition with varying load. Better classification is achieved with approach-2 using wavelet reconstructed signals. In the next section, the analysis is illustrated through a methodology with two approaches in brief, followed by results and discussion.

METHODOLOGY

The methodology presents the vibration analysis of gearbox fault diagnosis with two approaches. Figure 1 shows the schema involved in the fault diagnosis with two approaches (A1 and A2).

APPROACH-1

The sequence of steps involved in the manipulation of a raw vibration signal to diagnose the gearbox faults is shown in Figure 1-(a) as A_1. The approach started with partitioning the raw signal into segments/windows/epochs to reduce the complexity of huge data analysis and take advantage of data fusion. The signal length of each window has been converted into information theoretic discriminative values by computing statistical features. All the extracted features from each window are combined to form a feature set. Further, the concatenated feature matrix has been divided into training and testing data sets based on the requirements of the SVM fault classification algorithm. The final SVM classifier decision is an iterative result in terms of the mean (μ) and standard deviation (σ) of fault classification accuracy/success rate with 100 Monte Carlo runs.

APPROACH-2

The vibration analysis of gearbox fault diagnosis using DWT and wavelet reconstructed signal is shown in Figure 1-(b) as A_2. The raw vibration signal after windowing has been transformed into wavelet coefficients

reconstructed signals using DWT at different levels of decomposition. The remaining steps are the same as those of approach-1 with the addition of a feature selection block after feature extraction.

SENSOR DATA

The gearbox (wind turbine) vibration data has been recorded under varying load conditions from '0' to '90' percent using four accelerometer sensors mounted in four directions. The dataset includes healthy and broken tooth conditions with ten test cases for each condition. The resource of gearbox dataset has been obtained from the UCI Machine Learning Repository (Yogesh Pandya, 2018, Setti Suresh & VPS Naidu, 2018). Figure 2 illustrates the physical representation of a healthy signal and a broken tooth signal over a fixed window length of 4096 samples. It is observed that the amplitude of nonlinearity variation is less for broken tooth signal than a healthy signal.
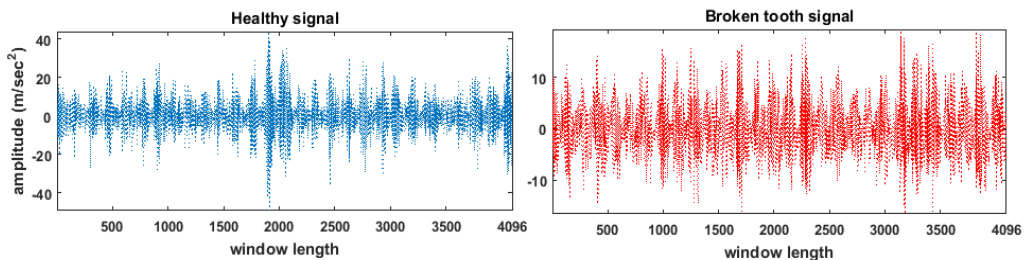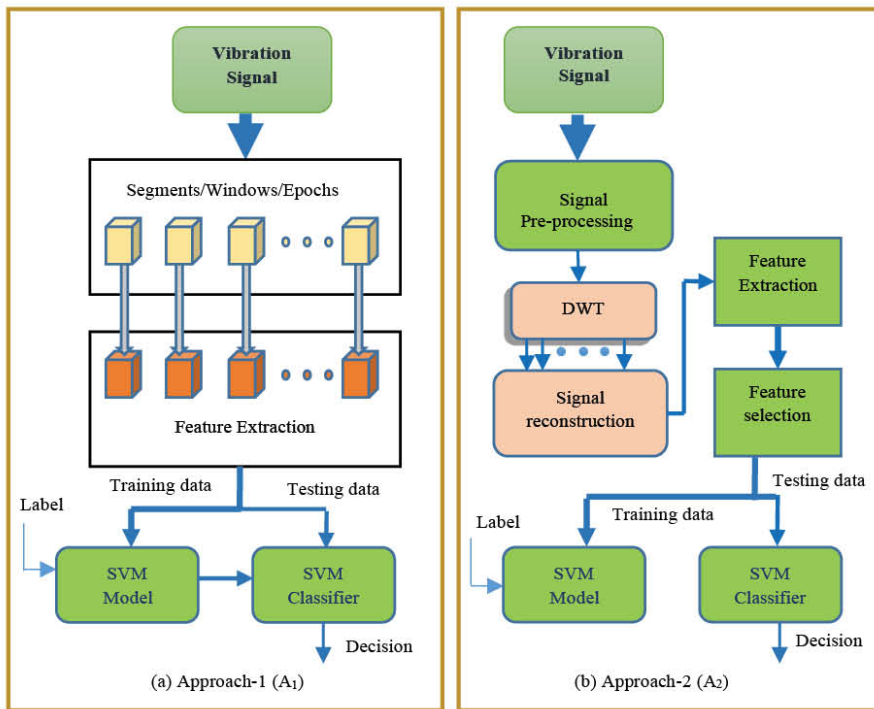
Figure 2. Illustration of gearbox healthy and broken tooth vibration signals.

DATA PREPROCESSING

The entire length of the gearbox vibration signal has been divided into segments/windows/epochs, giving more importance to data fusion. Each epoch is further manipulated according to the approach-1 and approach-2 to get the desired fault diagnosis results.

DISCRETE WAVELET TRANSFORM (DWT)

Discrete wavelet transform converts a discrete time series signal into one high-pass detailed coefficient series and one low-pass approximated coefficient series with suitable wavelet filters at each level of decomposition. For each epoch, wavelet transform has been applied with an increasing level of decomposition until the maximum fault classification accuracy is achieved. After decomposing the vibration signal into wavelet coefficients, the signal is reconstructed back from each wavelet coefficient, and further diagnostic features are extracted from the reconstructed signal.

FEATURE EXTRACTION

The raw vibration signal of a gearbox is very complex to set apart between healthy and broken tooth conditions. It necessitated performing feature extraction, which provides numerical-discriminative information to detect and diagnose the gearbox vibration signal. Table 1 shows the mathematical expressions for each estimated statistical feature from the wavelet coefficient's reconstructed signal. The statistical features used in this analysis are acquired from fault diagnostic research literatures (Wang, Xiang & Zheng, Yuan & Zhao, Zhenzhou & Wang, Jinping., 2015, Shanmukha Priya, V, Mahalakshmi P & VPS Naidu, 2015, Li, C., Sánchez, R.-V., Zurita, G., Cerrada, M., & Cabrera, D., 2016, Cemil Altın & Orhan Er, 2016, P. Vecer, M. Kreidl, R. Smd. 2005, & Setti Suresh & VPS Naidu, 2021). The features include mean (μ), absolute mean value (AMV), mean absolute deviation (MAD), standard deviation (σ), variance (σ^2), root mean square (X_{rms}), root value (X_{rv}), root sum of squares (X_{rssq}), peak-to-peak value (X_{p2p}), crest factor (X_{p2rms}), skewness value (X_{skew}), kurtosis value (X_{kurt}), skewness factor (X_{sf}), kurtosis factor (X_{kf}), margin factor (X_{mf}) or clearance factor (X_{clf}) or latitude factor (X_{lf}), energy (X_{energy}), and entropy (X_h). For each epoch (4096 samples), wavelet coefficient's reconstructed signal, we are getting 68 (3*17+17=68) features considering level-3 decomposition (3 detailed and 1 approximation coefficient reconstructed signal). Individual sensor observations of 483 are made by concatenating epochs for all 20 cases (10 healthy and 10 broken tooth) in rows. Feature matrix (483x69) was prepared by observations in rows and estimated features as variables in columns with load as a feature in the last column (both healthy and broken tooth observations have been combined for each individual sensor).

FEATURE SELECTION

The term feature selection has been used to get useful features from the estimated features ensuring increased fault classification accuracy. The well-defined infinite latent feature selection algorithm (ILFS) has been used as a feature selection algorithm (G. Roffo, S. Melzi, U. Castellani and A. Vinciarelli, 2017). Feature ranking was given from the estimated features, and top features were used with SVM classifier to increase the fault classification accuracy of the gearbox vibration signal.

SVM CLASSIFIER

Support vector machine, one of the superior classification and regression techniques that separates the feature space into subspaces through hyperplane attaining minimum error in classification, has been used in the diagnosis of the gearbox fault diagnosis. The feature matrix is divided into random permutated training data set with 70 percent

of observations and testing data set with remaining 30 percent of observations. The SVM model is trained with the training data, and fault classification accuracy of the trained SVM model is tested with a testing data using an built-in Matlab SVM classifier (Setti Suresh & VPS Naidu, 2021). The SVM algorithm was put under 100 MCR (Monte Carlo Runs) to eliminate the randomness in the success rate.

The mathematical expression satisfying the hyperplane of support vector machine is given by

$$y_i(wx_i + b) > 0.............y_i = \pm 1$$

$$y_i = label$$

$$x_i = feature\ vector$$

$$w = weight\ vector\ (orientation\ of\ the\ hyperplane)$$

$$b = bias\ term\ (position\ of\ the\ hyperplane)$$

RESULTS AND DISCUSSION

The vibration signal data of gearbox is tested with approach-1 and approach-2, and the accomplished results and observations made were discussed in this section. From the proposed comparative approaches in the fault diagnosis of the gearbox, Figure 3 shows the physical representation of the feature behavior for the wavelet coefficient's reconstructed signal at level-3 decomposition. All the estimated 17 features were graphically shown for detailed coefficients followed by the approximation coefficient. All the statistical features are computed for each wavelet coefficient reconstructed signal (F_1 to F_{17} are the statistical feature as stated in Table 1). It has been observed that 2–9, 14, 16, and 17 serial features (discussed in Features Extraction) are effective in providing discriminative information between healthy and broken tooth conditions with manual observation. However, no feature is providing selective information for the approximation coefficient reconstructed signal, indicating that no distinguished information exists in the lower frequency bandwidth.

Table 1. Statistical features.

S. No.	Features	S. No.	Features
1	$\mu = \dfrac{1}{N}\sum_{i=1}^{N} x_i$	9	$X_{p2p} = max(x) - min(x)$
2	$AMV = \dfrac{1}{N}\sum_{i=1}^{N} |x_i|$	10	$X_{p2rms} = \dfrac{max|x|}{X_{rms}}$

3	$$\text{MAD} \;=\; \frac{1}{N}\sum_{i=1}^{N}|x_i - \mu|$$	11	$$X_{skew} \;=\; \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3}{\left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}\right]^3}$$
4	$$\sigma \;=\; \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$$	12	$$X_{kurt} \;=\; \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4}{\left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}\right]^4}$$
5	$$\sigma^2 \;=\; \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$$	13	$$X_{sf} \;=\; \frac{X_{skew}}{X_{rms}^{\;3}}$$
6	$$X_{rms} \;=\; \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$$	14	$$X_{kf} \;=\; \frac{X_{kurt}}{X_{rms}^{\;4}}$$
7	$$X_{rv} \;=\; \left[\frac{1}{N}\sum_{i=1}^{N}|x_i|^{\frac{1}{2}}\right]^2$$	15	$$X_{mf} \text{ or } X_{clf} \text{ or } X_{lf} \;=\; \frac{\max|x|}{X_{rv}}$$
8	$$X_{rssq} \;=\; \sqrt{\sum_{i=1}^{N}x_i^2}$$	16	$$X_{energy} \;=\; \sum_{i=1}^{N}x_i^2$$
17	$$X_h \;=\; -\sum_{i=1}^{n}p(x_i)\log_2 p(x_i)$$		

x = wavelet coefficients reconstructed signal; i = index of   reconstructed signal from 1 to N; p_1, p_2…p_n = set of probabilities.

The SVM classification results for approach-1 and approach-2 with 100 Monte Carlo runs are shown in Table 2. The first column represents the label of approach-1 and approach-2 with varying level of wavelet decomposition. Analysis of individual sensors is performed to localize the fault and all sensors together to detect the fault. It can be noted that approach-2 with three levels of decomposition has achieved 100% fault classification rate for sensors 1 to 3 when compared with approach-1. But in the case of sensor 4, the SVM classifier could not converge by considering all the estimated features for classification. It motivated us to adopt a feature selection library (G. Roffo, S. Melzi, U. Castellani and A. Vinciarelli, 2017), and it was observed that, by using infinite latent feature selection (ILFS) algorithm, sorting the estimated top 20 wavelet reconstructed coefficient features in ranking driven to the classifier, the anomaly of no convergence was eliminated and led to a highest fault classification rate of 82.01% with a standard deviation of 2.96% (indicated with green color in the Table 2) when compared with approach-1 and success rate in (Setti Suresh & VPS Naidu, 2021). The execution time of the developed Matlab SVM algorithm has been examined

for A_1 and A_2. It is clear that approach-2 has less execution time up to sensor 3. For sensor 4, to get better classification accuracy/success rate, one has to compromise at execution time by adopting ILFS algorithm, which is shown in Table 2. Fault detection using all sensor features concatenation has the same classification accuracy for both approaches, indicating that the fault detection is simpler for this data. Individual sensor analysis was performed to localize the fault in practical scenario.
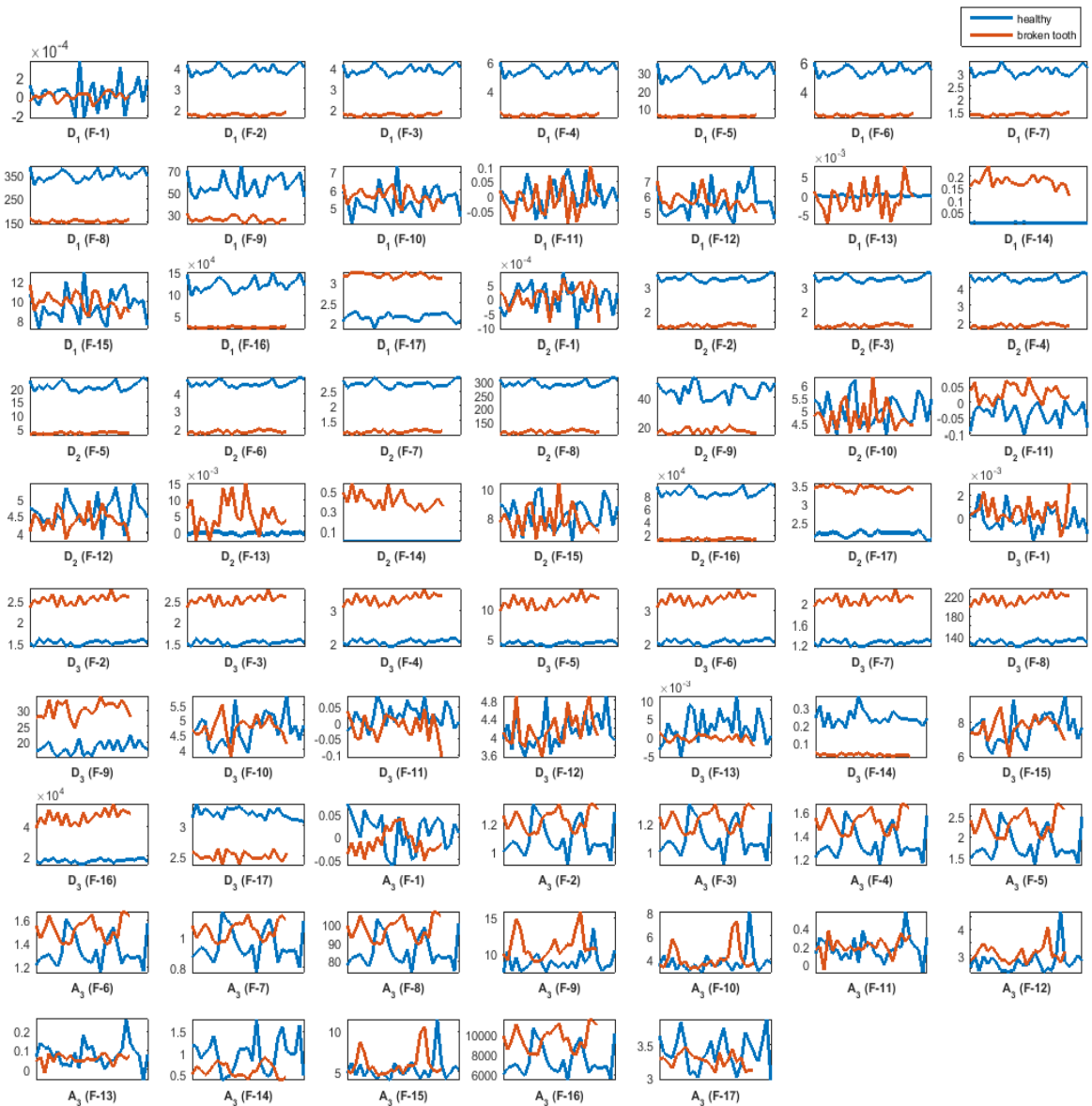
Figure 3. The physical representation of estimated feature behavior with wavelet coefficient's reconstructed signal (D1, D2, D3 = Detailed coefficient's reconstructed signal, A3 = Approximation coefficient's reconstructed signal).

The binary SVM fault classification plots using energy and entropy features for each sensor are shown in Figure 4. It can be inferred that sensor 1 is showing perfect classification, sensors 2 and 3 show moderate classification, and sensor 4 shows poor classification. From the numerical and graphical analysis, it is observed that the location of sensor 4 is not close to the location of the broken tooth gear mating, which resulted in poor classification.

Table 2. SVM classification results and execution time for A_1 and A_2 with 100 Monte Carlo Runs.

	Sensor 1		Sensor 2		Sensor 3		Sensor 4		All sensors
	SR	τ_c(ms)	SR	τ_c(ms)	SR	τ_c(ms)	SR	τ_c(ms)	
A_1	100 ± 0.00	104.41	98.32 ± 1.05	287.17	83.32 ± 2.62	437.69	69.19 ± 3.43	510.17	100 ± 0.00
A_2 L=1	100 ± 0.00	111.51	100 ± 0.00	104.16	99.95 ± 0.18	130.76	78.49 ± 3.10	843.05	100 ± 0.00
A_2 L=2	100 ± 0.00	106.90	100 ± 0.00	104.87	100 ± 0.00	134.02	NC/63.38 ± 3.73	NC/ 680.07	100 ± 0.00
A_2 L=3	100 ± 0.00	115.85	100 ± 0.00	111.88	100 ± 0.00	147.37	NC/82.01 ± 2.96	NC/ 983.32	100 ± 0.00
A_2 L=5	100 ± 0.00	117.37	100 ± 0.00	124.43	100 ± 0.00	164.50	NC/79.58 ± 3.03	NC/ 1071.78	100 ± 0.00
Setti. 2021	100 ± 0.00	-	100 ± 0.00	-	100 ± 0.00	-	81.85 ± 2.71	-	100 ± 0.00

A_1 = Approach-1; A_2 = Approach-2; SR = Success Rate (%); L = level of wavelet decomposition;
NC = No Convergence; τ_c = Elapsed SVM classification time.

From the above analysis of results and discussion, it has been noted that approach-2 (wavelet coefficient's reconstructed signal) is providing better classification accuracy of gearbox vibration signal when compared with approach-1 and methodology in [Setti Suresh & VPS Naidu, 2021] suitable for implementation in practical scenario.
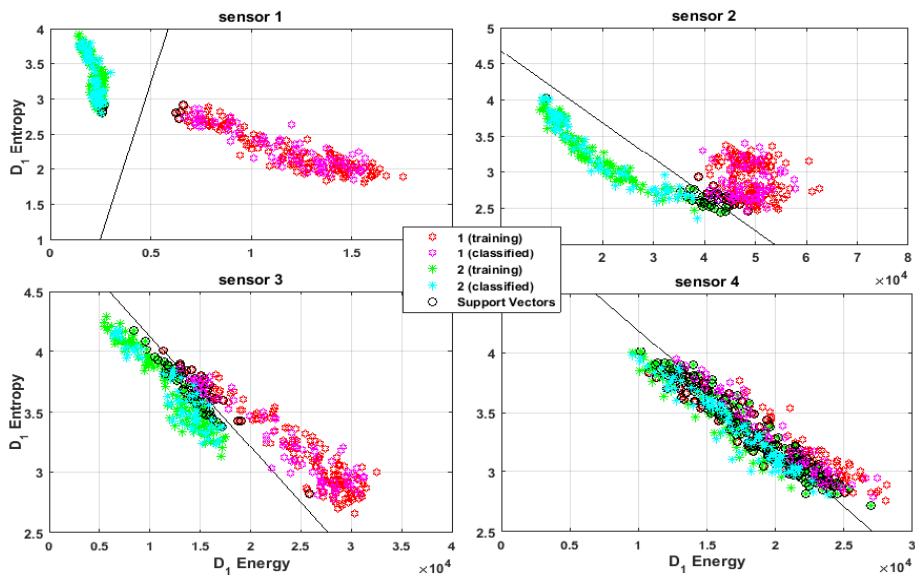
Figure 4. SVM fault classification using Energy and Entropy features for all sensors.

CONCLUSION

The analysis highlights the application of wavelet reconstructed signal in the enhancement of gear fault diagnosis. In the attempt to diagnose the gearbox faults, we adopted the wavelet reconstructed phenomena for comparison with traditional time-domain approach. The sequence of steps, segmentation, DWT, feature extraction, feature selection, and feature classification involved in the fault diagnosis of each approach have benefited the fault diagnosis. The random behavior of the estimated features with the wavelet coefficient's reconstructed signal is discussed, and observations are made through physical representation. With the concept of wavelet coefficients reconstructed signal and feature selection algorithm, the fault classification accuracy of sensor four in the dataset has improved, which is far away from the fault. The randomness in the fault classification is eliminated using 100 Monte Carlo Runs. Therefore, it is clear from the analysis that a more accurate maximum fault classification is achieved using wavelet reconstructed signal approach method than traditional time-domain approach.

REFERENCES

Wang, W. J., & McFadden, P. D., 1996. Application of Wavelets to Gearbox Vibration Signals for Fault Detection. Journal of Sound and Vibration, 192(5), 927–939.

Fei, S. 2017. Fault Diagnosis of Bearing Based on Wavelet Packet Transform-Phase Space Reconstruction-Singular Value Decomposition and SVM Classifier. Arabian Journal for Science and Engineering, 42(5), 1967–1975.

Cao Zhitong, Chen Hongping, He Guoguang, & Ritchie, E. (n.d.), 2001. Rotor fault diagnosis of induction motor based on wavelet reconstruction. ICEMS 2001. Proceedings of the Fifth International Conference on Electrical Machines and Systems (IEEE Cat. No.01EX501).

Yogesh Pandya, 2018. Retrieved from the data world: https://data.world/gearbox/gear-box-fault-diagnosis-data-set as on 23rd March 2021.

Setti Suresh & VPS Naidu, 2018. Gearbox Health Condition Monitoring: A brief exposition, Control and Data Fusion (CADF), e-Journal, Vol. 2, No. 4, pp. 13-23, Jul-Aug 2018.

Wang, Xiang & Zheng, Yuan & Zhao, Zhenzhou & Wang, Jinping., 2015. Bearing Fault Diagnosis Based on Statistical Locally Linear Embedding. *Sensors* (Basel, Switzerland). 15. 16225-47.

Shanmukha Priya, V, Mahalakshmi P & VPS Naidu, 2015. Bearing health condition monitoring: Wavelet decomposition. Indian Journal of Science and Technology, Vol. 8, Issue No 26, 2015.

Li, C., Sánchez, R.-V., Zurita, G., Cerrada, M., & Cabrera, D., 2016. Fault Diagnosis for Rotating Machinery Using Vibration Measurement Deep Statistical Feature Learning. Sensors, 16(6), 895.

Cemil Altın & Orhan Er, 2016. Comparison of Different Time and Frequency Domain Feature Extraction Methods on Elbow Gesture's EMG, European Journal of Interdisciplinary Studies Articles, European Center for Science Education and Research, vol. 5, May-August.

P. Vecer, M. Kreidl, R. Smd. 2005. Condition Indicators for Gearbox Condition Monitoring Systems, Acta Polytechnica Vol. 45 No. 6, 2005.

Suresh S., Naidu V.P.S. 2021 Gearbox Health Condition Monitoring Using DWT Features. In: Rao J.S., Arun Kumar V., Jana S. (eds) Proceedings of the 6th National Symposium on Rotor Dynamics. Lecture Notes in Mechanical Engineering. Springer, Singapore.

G. Roffo, S. Melzi, U. Castellani and A. Vinciarelli, 2017. Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach, IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 1407-1415.