

Condition monitoring of railway vehicles by Graph Signal Processing (GSP) of vibration signals

Dr. Jyoti Kumar Barman¹, Dr. Nupur Choudhury², Dr. Gitu Das³, Dr. Pushpanjalee Konwar⁴, Dr. Rupesh Mandal⁵ and Dr. Durlav Hazarika⁶

^{1,2,3,4,5} Assam Don Bosco University

¹ jyoti.barman@dbuniversity.ac.in

² nupur.choudhury@dbuniversity.ac.in

³ gitu.das@dbuniversity.ac.in

⁴ pushpanjalee.konwar@dbuniversity.ac.in

⁵ rupesh.mandal@dbuniversity.ac.in

⁶ Assam Engineering College

⁶ dlhazarika12@gmail.com

Abstract: Vibration analysis is one of the most popular and reliable methods for monitoring the health of various mechanical and civil structures, including rotating machines and railway vehicles. This paper presents a framework based on Graph Signal Processing (GSP) for condition monitoring of trains. Vibration signals are collected from sensors installed on tracks, capturing data as trains move over them. These signals are then represented on the nodes of a graph constructed from a database. The Graph Fourier Transform (GFT) along with a first order filter is employed for quantifying the conditions of the train under study.

Keywords: Vibration Data, Health Monitoring, Statistical Measures, Graph Signal Processing.

(Article history: Received 6th August 2026 and accepted 17th August 2026)

I. INTRODUCTION

Condition monitoring of railway vehicles is important in country like India, where a large number of trains ply daily. Periodical maintenance of railway vehicles can be streamlined if they are categorized based on their health conditions. Once categorized, maintenance schedules can be tailored to the specific needs of each train, ensuring optimal health of both the vehicles and the tracks. Proper maintenance significantly reduces the risk of accidents, especially derailments, which often result from poor track conditions or malfunctioning vehicle components such as wheels and suspension systems. Indian Railways runs more than 20,000 passenger trains every day. Monitoring the condition of all these trains manually in a laboratory setting is an extremely tedious task. Therefore, developing a monitoring system that can estimate the condition of the train while they are still running is indeed in demand. There are several systems available in this regard. Vibration analysis is a proven method for estimating the health of these components. Vibration signals generated by vehicle components can be analyzed to diagnose faults effectively [1], [2]. Traditional fault detection methods involve analyzing statistical indicators like Root Mean Square (RMS), Crest Factor (CF), and Kurtosis. These methods have been widely applied to detect faults in mechanical systems, particularly bearings [3], [4], [5]. While these statistical features are useful for tracking the degradation of a specific machine, comparing two systems based on these features can be challenging.

Condition monitoring, characterization, and classification of systems depend on the features they generate. Therefore, feature selection is crucial, and the selected features must be distinctive and sharp enough to facilitate easy classification or characterization. For similar systems, such as two trains, the differences in lower-order features like mean or RMS value are often minimal. Thus, improving these features is necessary to simplify the task of classifying or quantifying similar systems. Additionally, many researchers have used these features as inputs to artificial intelligence systems for condition monitoring [6], [7], [8]. Enhancing these features can expedite the learning process. Sensor fusion, which combines data from different sensors to improve features, has also been explored by several authors [9], [10]. In this paper, higher-order features such as skewness, kurtosis, and entropy, derived from probability distributions, are considered. However, because the vibration signals generated by moving trains are random and do not fit any specific probability distribution, calculating these higher-order features requires the use of the Empirical Cumulative Distribution Function (ECDF). Despite this, ECDFs derived from different vibration signals can appear quite similar, making system characterization difficult. Hence a Graph Signal Processing (GSP) based framework is proposed to categorize trains effectively. Further, it is found that correlating accelerometer and gyroscope signals enhance the statistical measures such as skewness and kurtosis.

II. METHODS

The methodologies of the work : At first, a database of railway vehicle vibrations developed by Neri et. al. is considered [11]. Based on this database, a graph is created to study particular vibration signals on it. Vibration signals of the trains of interest (Rajdhani Express, Kamrup Express and Kolongpar express) are collected. These signals are analyzed on the created graph for health estimation of the said trains. The flowchart of the process is shown in Fig. 1.

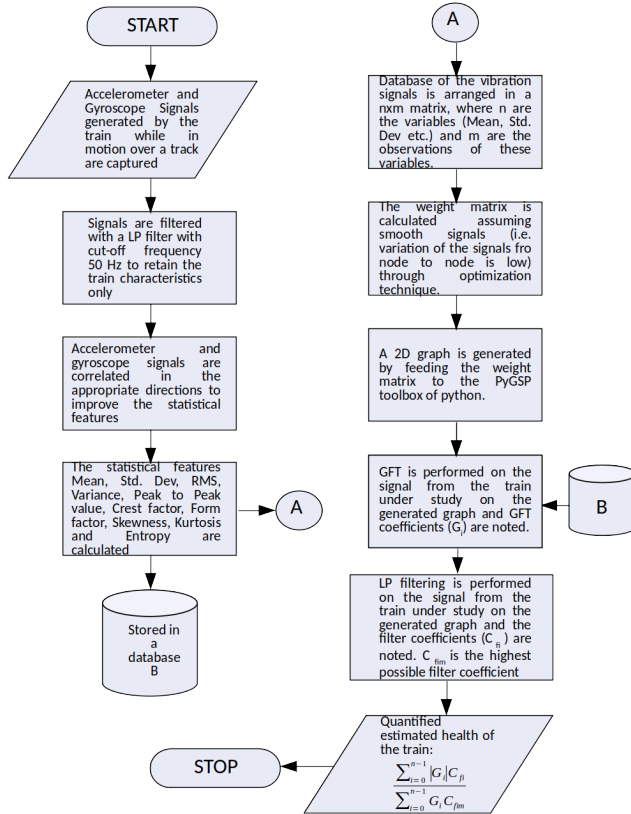


Fig. 1. Flowchart of the proposed method

A. About the database

The database created by Neri et. al. is available at <https://realvibrations.nipslab.org/> [11]. The database contains a lot of real world vibrations generated by different sources, such as transportation, human motions, devices, bridges etc. Seven trains that are available in this database are considered. From each train, 3 vibration signals in three axes viz. X,Y and Z-direction are available. Thus, a total of 21 vibration signals are being used to create the graph that can represent train vibration signals effectively.

B. Creating the graph

A graph can be learned from data. Given m observations on n variables, one data matrix $X \in R^{(n \times m)}$ can be created. Given some prior knowledge about the data; a graph G can be constructed from the data matrix X . Number of nodes of the graph is n and each column of X represents one signal on the graph [12]. To construct the graph, first thing is to

consider the variables or data entities, the classical ten statistical measures viz. mean, std. dev, RMS, variance, peak to peak value, crest factor, form factor, skewness, kurtosis and entropy are considered. These statistical measures will represent the $n=10$ nodes of the graph.

Given the variables and the observations as signals, a graph can be inferred in terms of Laplacian L . There is a one-to-one mapping between the Laplacian matrix L and the weighted undirected graph G and thus this is suitable to infer Laplacian L to learn the graph G . The data matrix $X \in R^{(n \times m)} = [x_1, x_2, x_3, \dots, x_n]^T$, where $x_1, x_2, x_3, \dots, x_n$ are the signals generated on the n nodes. The smoothness of the graph signal can be estimated by

$$\frac{1}{2} \sum_{i,j} w_{ij} \|x_i - x_j\|^2 \quad (1)$$

as $\|x_i - x_j\|^2$ represents the variation among the nodes and w_{ij} is a scaling factor. In fact w_{ij} denotes the weights of the edge between nodes i and j . The term $1/2$ is introduced to get a standard form so as to write in terms of Laplacian as

$$\frac{1}{2} \sum_{i,j} w_{ij} \|x_i - x_j\|^2 = \text{tr}(X^T L X) \quad (2)$$

Now the task of learning the graph is reduced to the optimization problem of the following form

$$\min_L \text{tr}(X^T L X) + f(L) \quad (3)$$

$f(L)$ is introduced to impose additional constraints to obtain a valid Laplacian L . Although the objective function in terms of the Laplacian has many advantages for reducing the search space for optimal solution, such as for an undirected graph the Laplacian matrix is symmetric; also, Laplacian is defined in such a way that the sum of each row (column) elements is zero, yet it is useful to minimize the problem in the following form [13]

$$\frac{1}{2} \sum_{i,j} w_{ij} \|x_i - x_j\|^2 = \text{tr}(W \circ Z) \quad (4)$$

where, $Z \in R^{(n \times n)}$ is the pairwise distance matrix given by $z_{ij} = \|x_i - x_j\|^2$ and " $\circ$ " represents hadamard product. Therefore, the task of learning the graph is reduced to the optimization problem

$$\min_W \text{tr}(W \circ Z) + f(W) \quad (5)$$

Searching for optimal solution $W \in \Omega$ is advantageous as few constraints can be included to reduce the search space such as $W \in R^{(n \times n)} : W = W^T$ and $\text{diag}(W) = 0$. Also since connecting a valid edge between two nodes is a problem of selecting two nodes out of n nodes, it can be represented by n choose 2, thus reducing the search space further [13]. $f(W)$ is introduced to impose further constraints to the minimization problem. One standard problem formulation including $f(W)$ in (5) is given by

$$\min_W \frac{1}{2} \text{tr}(W \circ Z) + 2\sigma^2 \sum_{i,j} w_{ij} (\log(w_{ij}) - 1) \quad (6)$$

Introducing the log term in the problem formulation prevents W from going to trivial solution of $W=0$. Again formulating the problem in this manner, a closed-form solution can be obtained which is

$$w_{ij} = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}} \quad (7)$$

a Gaussian distribution. And thus the required weight matrix for the intended graph learning can easily be calculated with (7) by simply considering the individual columns of the data matrix X .

C. Data collection for the trains of interest

In this study three trains are considered, viz. Rajdhani Express, Kamrup Express and Kolongpar Express. These three trains were suggested by NFR authority to carry out the experiments as they have different range of travel distances with different speeds and they come under different category of trains according to NFR. The Indian railway tracks are fixed on precast RCC Sleeper using fishplate. This type of railway track not only vibrates when a train moves over it, the angular motion of the track is also experienced. Therefore, it is necessary to capture both vibrations and angular velocities of a track for proper health estimation of the trains. The work reported in this paper, an MPU6050 was used to capture the acceleration and the angular velocity of a railway track while a train in motion. Arduino Uno development board was used to interface MPU6050 and a laptop through the USB port for capturing MPU6050 signals while a train was in motion. The acceleration and the angular velocity of a train were analyzed. The angular velocity from the gyroscope of the MPU6050 and the vibration signal from the accelerometer of the MPU6050 (represented as velocity derived from acceleration) were correlated. Analyses were carried out to find the proper combination of signals that to be correlated for better extraction of the features of a train under motion. The time-domain features viz. skewness, RMS, kurtosis, mean, standard deviation, variance, crest factor, form factor, peak to peak value and entropy of the accelerometer signal, the gyroscope signal and the correlated signals were calculated.

A simple Arduino based system has been configured to capture the vibration and the angular velocity of a train, which is generated when the train moves over a railway track. The sensor MPU6050 is interfaced to an Arduino Uno card to capture vibration and the angular velocity of a railway track. The MPU6050 sensor was placed on the wayside of the track's fishplate to ensure consistent performance and protect it from possible impact due to environmental changes. MPU6050 is an integrated 6-axis motion tracking device that combines a 3-axis gyroscope, 3-axis accelerometer along with a Digital Motion Processor. It has in-built analogue-to-digital converters (ADC), with operating software that takes care signal processing of the captured signals and to deliver digital output signals with lowest possible noise. The sensor is one of the most reliable one as it passes through a number of tests before its release. Moreover its robust nature to noise and temperature stability makes the sensor very desirable. For capturing the signals by Arduino Uno card, a library programme developed by Mr. Jeff Rowberg is used. This library manipulates the Digital Motion Processor (DMP) and provides "Real world Acceleration" along with other data values. Combination of this programme and the lowest noise offered by the MPU6050 eliminates traditional difficulties of sensing such as noise, extracting the true values from the raw data etc. Arduino based system has a software platform that allows

incorporation of in-built software prototype for data acquisition from MPU6050 sensor and store the data (vibration and the angular velocity) in the Laptop. The function block diagram of the measurement scheme is shown in Fig. 2(a). Fig. 2(b) shows the orientation of the sensor MPU6050, when placed on the track.

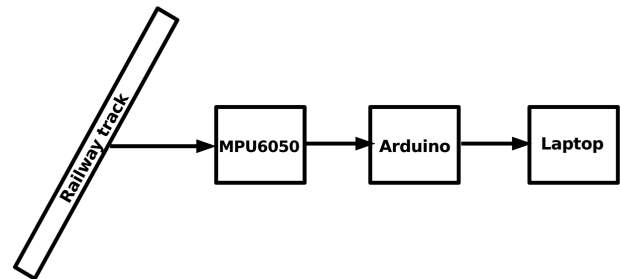


Fig. 2(a). Schematic of the data acquisition system

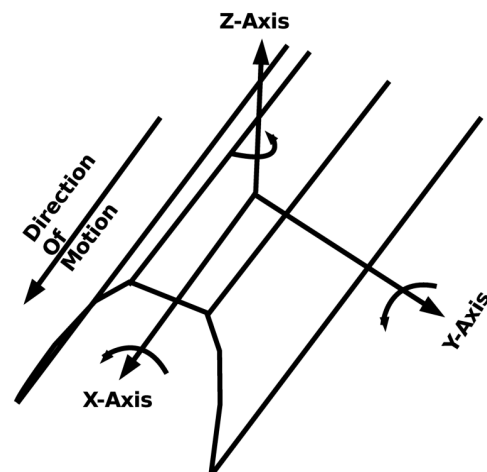


Fig. 2(b). Orientation of the sensor MPU6050 as placed on the railway track.

D. Frequency domain analysis of GSP

To compare the health of the different trains, frequency domain analysis of GSP is utilized. Graph Fourier Transform (GFT) is utilized for this purpose. Given a graph G and a graph signal $x \in \mathbb{R}^n$ defined on G ; a normal Graph-Shift operator $S = V \Lambda V$ can be considered. Columns of $V = [v_1 v_2 v_3 \dots v_n]$ corresponds to the eigenvectors of S . Λ is a diagonal matrix containing the eigenvalues of S . Now, GFT of x can be defined as

$$\tilde{x}(k) = \langle x, v_k \rangle = \sum_l x(l) v_k^*(l) \quad (8)$$

Where $*$ represents conjugate. While learning the graph, it was stressed on the fact that the signals on the graph are smooth. The eigenvalues of the graph are equivalent to the frequency values in Fourier Transform and the eigenvectors corresponding to the eigenvalues represent different orthogonal directions in the space. Therefore, a low frequency (i.e. lower eigenvalues) represent a lesser

variation of the signals on the graph. A low pass filter of the form $1/(1 + \tau x)$ along with the GFT components can be suitably utilized for the quantification of train health. We propose the following formula for health estimation of the trains based on GFT and low pass filter coefficients.

$$EstimatedHealth = \frac{\sum |G_i| C_{fi}}{\sum G_i C_{fim}} \quad (9)$$

where, G_i is the GFT value corresponding to the eigenvalue i , C_{fi} is the low pass filter coefficient of the eigenvalue i . C_{fim} is the highest possible filter coefficient for the low pass filter considered.

Frequency response of the filter $1/(1 + \tau x)$ is shown in Fig. 3. In (9), while in the numerator, absolute values of GFT are considered, in denominator algebraic values of GFT are considered. It is to make sure that $||Estimated\ health|| \leq 1$

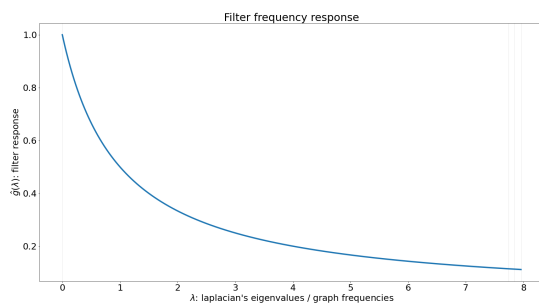


Fig. 3. Response of the filter $1 / (1 + \tau x)$

III. RESULTS AND ANALYSIS

A. Creating the graph

The reference dataset provided by [11] incorporates 3-axis vibration records for 7 distinct locomotives, which yields 21 individual columns for our target data matrix. Table I displays a representative segment of this matrix. The requisite weight matrix is derived utilizing (7), ensuring that all diagonal values are set to zero. Subsequently, this calculated matrix is processed using the Python-based PyGSP library to construct the final graph structure [14]. The resulting network is illustrated in Fig. 4.

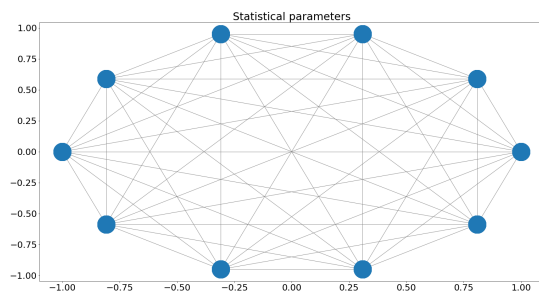


Fig. 4. Graph created from the database [11]

Table I: Snippet of the data matrix from database [11] used for the creation of the graph

Sl. No.	Statistical Measures	Train 1-X	Train 1-Y	Train 1-Z	Train 2-X	Train 2-Y	Train 2-Z	Train 3-X	Train 3-Y	Train 3-Z
1	Mean	-0.00014	0.00755	0.00135	0.00108	0.00032	0.00163	-0.00052	0.00068	0.000749
2	Std. Dev	0.04659	0.03279	0.009467	0.0029	0.002637	0.006	0.001374	0.001129	0.002
3	RMS	0.04659	0.03364	0.009468	0.0029	0.00269	0.006	0.001375	0.0013	0.002
4	Variance	0.00217	0.00107	0.000896	0.000084	0.000069	0.000369	0.000018	0.000012	0.000041
5	X_pp	10.66	9.3895	16.28	0.74522	0.3127	1.75	4.5	3.854	8.14
6	CF	129.14	154.62	74.15	12.55	5.6807	13.66	199.52	153.02	173.29
7	FF	-33.22	4.45	69.86	26.93	5.05	37.22	-26.34	1.99	11.68
8	Sk	1.83143	6.87	-1.04184	0.0079	0.004555	-0.005	12.93	1.26	-13.73
9	Ku	714.22	2631.4	355.88	3.89	0.49859	5.69	5459.96	5340.04	10815.2
10	En	-78.71	63.6248	62.433	1056.04	4844.86	1168.96	-793.64	7716.52	1887.44

B. Time-Domain Features

Fundamental time-domain metrics---including mean, variance, standard deviation, and RMS---derived sequentially from the MPU6050 sensor's vibratory and angular velocity outputs act as baseline indicators of a vehicle's mechanical integrity. Consequently, logging these variables offers substantial utility for tracking a specific train's condition across extended operational cycles. Additionally, these parameters facilitate real-time qualitative assessments. By archiving both the raw vibration signatures and these derived low-order metrics, operators can monitor progressive mechanical wear; for instance, a steady climb in RMS often points to compounding structural defects [15]. Certain temporal features, namely crest factor and form factor, respond primarily to signal geometry rather than amplitude. This characteristic makes them highly suitable for evaluating parallel systems. Nonetheless, their volatility in response to minor fluctuations in baseline metrics (like mean or RMS) limits their reliability for rigorous diagnostic evaluations. Because individual trains possess unique design architectures, velocity profiles, and operating conditions, their resultant vibration waveforms naturally distort, subsequently altering these shape-dependent features. To bypass these constraints, advanced higher-order statistical metrics---skewness, kurtosis, and entropy---are implemented. Stemming from the mathematical moments

that define the signal's probability density function (PDF), these parameters remain largely insulated against external influences such as train speed or physical construction.

1. Definitions and mathematical formulae of the time-domain features considered

Assuming $x(n)$ represents a continuous signal sequence, the specific statistical parameters evaluated in this study, alongside their governing equations, are detailed in table II.

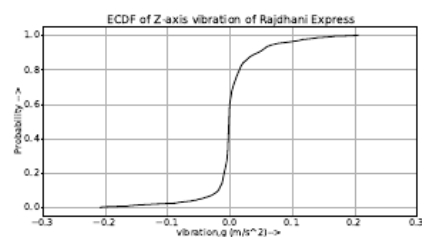
Table II: Time-domain features

Sl No	Features	Formula	Remarks
1	Mean	$\mu = (\sum x(n))/N$	$n=1,2,3,...,N$ are the data points
2	Std. Dev	$\sigma = \sqrt{\sum (x(n)-\mu)^2 / (N-1)}$	$n=1,2,3,...,N$ are the data points [16]
3	RMS	$X_{rms} = \sqrt{\sum x(n)^2 / N}$	$n=1,2,3,...,N$ are the data points [16], [17]
4	Variance	$v = \sum (x(n)-\mu)^2 / (N-1)$	
5	Peak to Peak value	$\max(x(n)) - \min(x(n))$	
6	Crest Factor	Positive Peak / RMS [18]	Standard deviation divided by RMS value [17].
7	Form Factor	$SF = \sqrt{((1/N)\sum x(n)^2) / ((1/N)\sum x(n))}$	Modification is done in calculating mean by considering the absolute values of the signal sequence. [19]
8	Skewness	$Sk = m_3 - 3m_2m_1 + 2m_1^3$	m_1 is the first moment coefficient, m_2 is second, m_3 is third [20].
9	Kurtosis	$Ku = m_4 - 3m_2^2 - 4m_3m_1 + 12m_2m_1^2 - 6m_1^4$	where m_1 to m_4 are moment coefficients [20].
10	Entropy	$E_{Sh} = \sum x_i \log(1/x_i)$	Shannon entropy is considered [21]

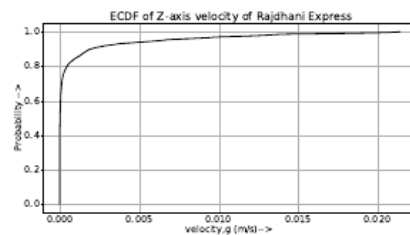
2. Calculation of the statistical features of the trains of interest

Testing protocols were strictly governed by the NFR administration, which constrained empirical data collection to a specific timeframe and three distinct rail services: the Kolongpar Express, Rajdhani Express, and Kamrup Express. Security and administrative personnel continuously monitored all field experiments. The Kamrup Express functions as a vital transport corridor linking Dibrugarh, Assam, with Howrah, West Bengal. Functioning as a flagship service for the NFR, the Rajdhani Express (Dibrugarh) maintains an average velocity of 65 kmph over a 2434 km journey. Meanwhile, the Kolongpar Express provides connectivity between Mairabari and Guwahati. The NFR specifically recommended these three locomotives for evaluation because they encompass a broad spectrum of speeds, operational distances, and official administrative classifications.

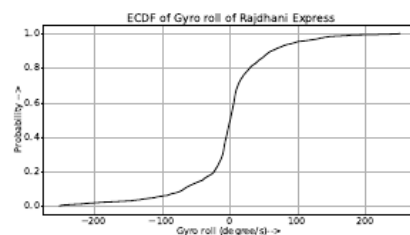
During transit, six independent data streams were recorded to a mobile workstation for each train: the X, Y, and Z spatial axes for physical acceleration (vibration) alongside the corresponding X, Y, and Z axes for angular velocity (gyroscope). Prior to in-depth evaluation, the raw acceleration telemetry generated by the MPU6050 was mathematically integrated into velocity vectors. This transformation was essential to enable accurate synchronization and cross-correlation with the gyroscopic feeds. Furthermore, determining specific higher-order traits---entropy, kurtosis, and skewness---required the computation of the Empirical Cumulative Distribution Function (ECDF). To illustrate this process, Fig. 5 displays the ECDF visualizations for the Rajdhani Express, plotting the Z-axis acceleration, Z-axis velocity, gyroscope roll angle, and a fused signal combining the gyro roll with the Z-axis velocity.



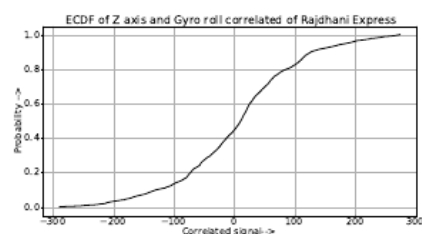
(a) ECDF of the Z-axis vibration (acceleration) of the Rajdhani Express



(b) ECDF of the Z-axis vibration (velocity) of the Rajdhani Express



(c) ECDF of the Gyro roll of the Rajdhani Express.



(d) ECDF of the Z-axis vibration and Gyro roll correlated signal of the Rajdhani Express.

Fig. 5. ECDFs of Rajdhani Express (a, b, c, d)

Another important aspect at this point was to select the axes to be correlated for proper meaningful results. As there are 6 different signals, 15 correlations would have been possible, but might not yield meaningful results. This issue is discussed in the following subsection. For correlation, both the signals must be of the same dimension. Therefore it was necessary to convert the acceleration signal of the accelerometer to velocity such that it could be correlated with the angular velocity of the gyroscope. Trapezoidal rule was utilized for numerical integration to obtain velocity from the acceleration samples.

3. Selecting proper signals for correlation

The primary objective of cross-referencing these disparate data streams is to accurately gauge the operational health of the rolling stock. Evaluating vibratory responses along the Z-axis provides critical insights into both overall mechanical condition and passenger ride quality, whereas examining the X and Y axes primarily highlights derailment risks and safety parameters [22]. The spatial configuration of the gyroscope, specifically when the accelerometer's X-axis is aligned horizontally through the elliptical center, is depicted in Fig. 6. A review of this orientation confirms that merging the gyroscope's roll data with the accelerometer's Z-axis output yields a highly detailed diagnostic view of vertical train vibrations. Applying this same geometrical logic, optimal signal resolution is achieved by pairing the gyroscope's pitch with the Y-axis, and the gyroscope's yaw with the X-axis.

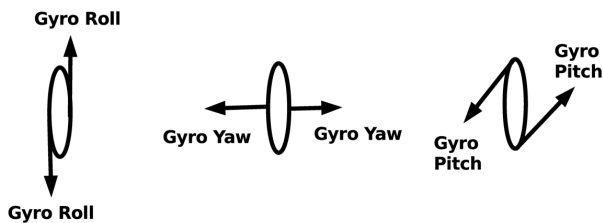


Fig. 6. Gyroscope orientation

Comprehensive statistical profiling---encompassing mean, standard deviation, RMS, variance, peak-to-peak amplitude, crest factor, form factor, skewness, kurtosis, and entropy---was executed across the raw gyroscopic streams, the mathematically derived velocity streams, and the synthesized correlated signals. The Rajdhani Express was analyzed initially, with its processed metrics documented in table III. Following identical procedures, the equivalent time-domain parameters for the Kamrup and Kolangpar Express services were extracted and are organized in tables IV and V, respectively.

Table III: Time-domain features of Rajdhani Express

Sl No	Fe at ur es	Velocity (m/s)			Angular velocity (deg/s)			Correlated		
		X-	Y-	Z-	R	Y	Pit	X-	Y-	Z-

		A x	A x	A x	oll	a w	ch	A x and Y a w	A x and Pit ch	A x and R oll
1	Mean	0.0009	0.0007	0.0048	-0.765	-1.1294	7.31	-12.743	27.508	20.681
2	Std. Dev	0.0029	0.0019	0.0072	62.977	54.291	49.484	83.288	52.324	19.7099
3	RMS	0.003	0.002	0.0087	62.982	54.303	50.021	84.258	59.115	19.818
4	Variance	8.7249	3.5465	0.0005	39.661	29.475	24.487	69.370	27.378	38.841
5	X _{pp}	0.0214	0.0198	0.0444	50.1504	51.399	49.098	64.882	40.142	10.1899
6	CF	6.9203	9.9341	5.0524	3.9785	4.8323	5.0013	2.7896	5.0322	2.687
7	FF	3.4013	2.9494	1.7927	-82.322	-48.081	6.8428	-6.6117	2.1489	9.5825
8	Sk	4.4484	6.0905	2.2152	-0.4508	0.0229	0.9777	-0.7259	1.5668	0.1383
9	Ku	21.405	46.969	5.3917	4.0426	6.4322	8.1593	3.0875	4.3367	-0.2555
10	En	11.465	10.30	56.5	16.7313	83.926	-91.061	16.31005	-31.91688	-31.15827

Table IV: Time-domain features of Kamrup Express

Sl No	Fe at ur es	Velocity (m/s)			Angular velocity (deg/s)			Correlated		
		X- A x	Y- A x	Z- A x	R oll	Y a w	Pit ch	X- A x and Y a	Y- A x and Pit ch	Z- A x and R oll

1	Mean	-0.0001	0.0003	-0.0031	5.71	-1.6911	2.6444	-2.8541	4.1414	-75.187
2	Std. Dev	0.0026	0.0029	0.0048	53.467	48.061	36.866	59.827	63.719	13.053
3	RM S	0.0026	0.003	0.0058	53.771	48.091	36.961	59.895	63.854	15.0642
4	Variance	6.8583	8.9547	2.3518	2858.73	2309.89	1359.2	3579.7	4060.2	17.0404
5	X _p	0.0209	0.0211	0.0265	50.45	50.278	36.027	49.472	52.899	81.416
6	C F	7.6625	6.8223	3.44	4.67	5.2529	6.7682	4.1086	6.0423	1.7849
7	FF	-22.487	8.2877	-1.821	9.4159	-28.436	13.976	-20.984	15.418	-2.0035
8	Sk	5.537	5.0938	2.3294	0.6824	-0.2135	3.1853	-0.2454	2.4487	-0.249
9	K u	33.105	26.706	5.14	6.92	9.4051	16.575	3.5454	9.6257	0.453
10	En	-5.5541	1.8131	-45.16	-79.738	23.253	-43.708	41.05	-89.812	10.3551

Table V: Time-domain features of Kolongpar Express

Sl No	Features	Velocity (m/s)			Angular velocity (deg/s)			Correlated		
		X-Axis	Y-Axis	Z-Axis	Roll	Yaw	Pitch	X-Axis and Yaw	Y-Axis and Pitch	Z-Axis and Roll
1	Mean	0.0001	0.0001	0.0017	-1.624	1.5	1.8156	2.3672	0.128	-12.7015

2	Std. Dev	0.0014	0.0002	0.0035	31.1843	30.2434	14.8615	16.7497	1.6571	45.9526
3	R M S	0.0014	0.0002	0.0039	31.2266	30.2805	14.972	16.9162	1.6621	47.6756
4	Variance	2.0418	0.0000	0.0001	972.4462	914.6634	22.0861	28.05539	2.7461	21.11643
5	X _p	0.0211	0.0034	0.0246	44.6193	37.42255	21.81555	18.878	17.052	39.09628
6	C F	14.6223	11.6271	6.0888	8.0087	6.9242	9.3996	7.43	6.4054	3.9636
7	FF	7.58	27.6158	2.2817	-19.227	20.1832	8.2462	7.1458	12.9827	-3.7535
8	Sk	10.2259	5.0938	3.8192	1.1052	1.3124	3.1853	2.5512	2.4487	0.2078
9	K u	12.09	64.9	15.0694	23.7358	15.8028	35.925	16.47	15.776	4.5188
10	En	2.1415	-0.1440	24.474	88.681	-21.161	-14.744	-25.925	-10.77	11.54421

C. GFT of the signals of the trains on the graph

Executing a Graph Fourier Transform involves mapping a specific signal array---representing a single data matrix column for a given locomotive---onto the master reference graph. Reviewing the Rajdhani Express data inside table III, nine unique signal columns are available for evaluation. Because the core objective is to benchmark mechanical health, analysts can isolate distinct signals based on specific diagnostic requirements. For instance, focusing on the Z-axis provides data on passenger comfort, whereas safety assessments heavily rely on X and Y axis vibrations [22]. To establish an effective comparative baseline, it is advantageous to select the signal column that exhibits the most stable statistical profile. Advanced metrics like entropy, skewness, and kurtosis serve as excellent comparative tools for relative health indexing. Under theoretical flawless conditions, a mechanically sound system will output a skewness of 0 and a kurtosis of 3, while entropy tracks the inherent randomness within the signal [17].

An analysis of table III indicates that the X-Ax and Yaw column presents the most optimized metrics for the Rajdhani Express, achieving a kurtosis of 3.0875 and a skewness of -0.7259. This trend is mirrored in the Kamrup Express data (table IV). While the Kolongpar Express (table V) registered its most favorable metrics within the Z-Ax and Roll column, standardizing the evaluation dictated that all three vehicles be benchmarked using the X-Ax and Yaw data stream. It is important to acknowledge that this methodology allows for comparative analysis along whichever signal axis best serves the immediate diagnostic need. Finally, applying (9) computes the standardized health scores for these trains, which necessitates extracting the relevant GFT and low pass filter coefficients. These final quantifications are presented in table VI.

Table VI: Estimated health

Sl No	Eigen value	LP filter coefficient	GFT Coefficient		
			Rajdhani	Kamrup	Kolongpar
1	0	1.0	-2579.733	-1392.427	-179.652
2	1	1.0	16310.05	41030.5	-25923.25
3	2	0.9999999968	-908.799	-488.537	-46.053
4	3	0.0000591622	-671.365	-380.779	-40.643
5	4	0.0000525445	1441.02	780.568	84.544
6	5	0.0000454501	-1593.642	-837.67	-76.246
7	6	0.0000454372	-4853.833	-2502.075	-194.512
8	7	0.0000454367	-3446.882	-1762.645	-126.397
9	8	0.0000454363	22.508	11.804	0.724714181
10	9	0.0000453999	-754.079	-250.256	111.728
Estimated health =			0.89	0.79	-0.98

IV. DISCUSSION

In this paper, two major points are highlighted. Correlation of accelerometer and gyroscope signals improve the higher order statistical features: skewness and kurtosis in certain direction. This point can be verified experimentally and

evident from the tables III, IV and V. For ex, if in table III, if we compare the individual signals against the correlated signal of X-Ax and Yaw of the Rajdhani express, skewness and kurtosis improves drastically. As mentioned earlier, a healthy machine should produce kurtosis of 3 and skewness of 0 in ideal conditions. And we can see kurtosis of 3.0875 and skewness of -0.7259 for the said column of X-Ax and Yaw. Similar results can be readily seen in the table IV in the column of X-Ax and Yaw for Kamrup express and in table V in the column of Z-Ax and Roll for Kolongpar express.

The other major point is derivation of a framework using GSP to estimate the health condition of trains. From table VI, in this framework Rajdhani scores 0.89, Kamrup scores 0.79 and Kolongpar scores -0.98. The framework provides a way to quantify the health of a certain train. Since in this framework the maximum and minimum values are 1 and -1 respectively, we can say that the Rajdhani is the best performing while Kolongpar is the worst performing among the trains that have been compared. This agrees with the crude classification of trains by NFR, as it is mentioned that Rajdhani is a premier category train of NFR, while Kamrup is an ISO certified train but there is neither such certification nor any mention of being premier train for Kolongpar express.

V. CONCLUSION

The work reported in this paper presents the following findings: Correlation of accelerometer and gyroscope signals from a single sensor MPU6050 produces a better value for higher order statistical parameters. A framework is described to quantify the train health in the range of 1 to -1. The framework involves creating a graph from database, utilizing GFT coefficients and low pass filter coefficients of the vibration signal (from the train) to estimate the health of the train. A graph can be created with a vibration database with the statistical parameters as its nodes. Moreover, similar vibration signals can be processed on the created graph. In this work it is shown that by correlation, the kurtosis and skewness values can be pushed to their ideal values 3 and 0 respectively in certain directions. A graph is created with a train vibration database, nodes being Mean, Std. Dev, RMS, Variance, peak to peak value, crest factor, form factor, skewness, kurtosis and Entropy. Health of three trains Rajdhani, Kamrup and Kolongpar express have been evaluated with the said framework. It is found that Rajdhani scores 0.89, Kamrup scores 0.79 and Kolongpar scores -0.98 in this framework. 1 being maximum and -1 being minimum value for the output of the framework, it can be confirmed that Rajdhani enjoys better health compared to Kamrup, while Kolongpar is not in good condition.

REFERENCES

- [1] R. Yan, R. X. Gao, and C. Wang, "Experimental evaluation of a unified time-scale-frequency technique for bearing defect feature extraction," *J. Vib. Acoust.*, vol. 131, June 2009.
- [2] R. Yan and R. X. Gao, "Rotary machine health diagnosis based on empirical mode decomposition," *J. Vib. Acoust.*, vol. 130, February 2008.
- [3] N. Tandon, "A comparison of some vibration parameters for the condition monitoring of rolling element bearings," *Measurement*, 1994.

- [4] R. B. W. Heng and M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition," *Appl. Acoust.*, 1998.
- [5] H. Martin and F. Honarvar, "Application of statistical moments to bearing failure detection," *Appl. Acoust.*, 1995.
- [6] I. Abu-Mahfouz, "Drilling wear detection and classification using vibration signals and artificial neural network," *Int. J. Mach. Tools Manuf.*, 2003.
- [7] B. Samanta and K. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mech. Syst. Signal Process.*, 2003.
- [8] B. Sreejith, A. Verma, and A. Srividya, "Fault diagnosis of rolling element bearing using time-domain features and neural networks," in *2008 IEEE Region 10 Colloq. and 3rd Int. Conf. Ind. Inf. Syst. (ICIIS)*, Kharagpur, India, December 8–10, 2008.
- [9] K. M. Reichard, M. V. Dyke, and K. Maynard, "Application of sensor fusion and signal classification techniques in a distributed machinery condition monitoring system," in *Proc. SPIE*, 2000.
- [10] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, 2005.
- [11] I. Neri, F. Travasso, R. Mincigrucci, H. Vocca, F. Orfei, and L. Gammaitoni, "A real vibration database for kinetic energy harvesting application," *J. Intell. Mater. Syst. Struct.*, vol. 23, no. 18, pp. 2095–2101, 2012.
- [12] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 44–63, May 2019.
- [13] V. Kalofolias, "How to learn a graph from smooth signals," in *Proc. 19th Int. Conf. Artif. Intell. Statist. (AISTATS)*, A. Gretton and C. C. Robert, Eds., Cadiz, Spain, May 2016, vol. 51, pp. 920–929.
- [14] M. Defferrard, L. Martin, R. Pena, and N. Perraudin, "PyGSP: Graph signal processing in Python," 2017, <https://github.com/epfl-lts2/pygsp>.
- [15] Z. Shen, Z. He, X. Chen, C. Sun, and Z. Liu, "A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time," *Sensors*, 2012.
- [16] Y. Zhang, H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement*, vol. 46, pp. 1178–1189, 2013.
- [17] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*, 2017.
- [18] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," *Int. J. Adv. Manuf. Technol.*, 2017.
- [19] C. Yiakopoulos, K. Gryllias, and I. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a k-means clustering approach," *Expert Syst. Appl.*, vol. 38, pp. 2888–2911, 2011.
- [20] B.-S. Yang and K. J. Kim, "Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals," *Mech. Syst. Signal Process.*, vol. 20, pp. 403–420, 2006.
- [21] V. Sharma and A. Parey, "Gearbox fault diagnosis using RMS based probability density function and entropy measures for fluctuating speed conditions," *Struct. Health Monit.*, 2016.
- [22] J. K. Barman and D. Hazarika, "Condition monitoring of NFR trains with measurements from a single wayside 3D vibration sensor," *IEEE Sensors J.*, vol. 20, pp. 4096–4103, April 2019.

AUTHOR PROFILE



Jyoti Kumar Barman is an assistant professor in the department of Electrical and Electronics Engineering at Assam Don Bosco University. He holds a Ph.D. from Assam Engineering College under Gauhati University in the area of Electrical Engineering which was awarded in 2021.



He received his M.Tech degree from Assam Don Bosco University in the field of Digital Signal Processing in the year of 2015. He received his BE degree from Assam Engineering College in the field of Instrumentation Engineering in the year of 2009. His research interests includes condition monitoring, fault detection and identification, sensors, signal processing etc.

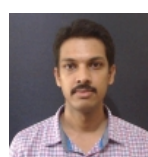
Nupur Choudhury is an Assistant Professor in the Department of Computer Science and Engineering at Assam Don Bosco University. She received her Ph.D. from Assam Don Bosco University in the area of Computer Science and Engineering in 2024. She received her M.Tech degree from Sikkim Manipal Institute of Technology in the field of Robotics in the year of 2015. Her research interests includes Signal Processing condition monitoring, IoT, sensors, etc.



Gitu Das is an assistant professor in the Department of Electrical and Electronics Engineering at Assam Don Bosco University, Assam. He has earned his PhD from Assam Engineering College under Gauhati University in the area of Electrical Engineering in 2024. He received his M.Tech degree from Assam Don Bosco University in the field of Embedded Systems and his BE degree from MVJ College of Engineering, Bangalore in Electrical and Electronics Engineering discipline. His research interests include power systems, embedded systems, IoT, signal processing, etc



Pushpanjalee Konwar is an Assistant Professor in the Department of Electrical and Electronics Engineering at Assam Don Bosco University, earned her Ph.D. in Electrical Engineering from NIT Nagaland in 2023. Her research interests include power system network reconfiguration, power system planning and operation, distribution system analysis, graph theory applications, Fault detection and identification etc.



Rupesh Mandal is an Assistant Professor in the Department of Computer Science and Engineering at Assam Don Bosco University. He has earned his Ph.D. from Assam Don Bosco University in the area of Computer Science and Engineering in 2026. He received his M.Tech degree from

Sikkim Manipal Institute of Technology in the field of Robotics in the year of 2016. His research interests includes, IoT, sensors, Fault Detection and Identification etc.



Prof. Durlav Hazarika obtained his Bachelor of Engineering in the field of Electrical Engineering in the year 1983 from Jorhat Engineering College. He completed his Master degree in Electrical Engineering (Power System and Apparatus) from IIT, Bombay in the year 1986. He obtained his Ph.D. degree from IIT, Kharagpur in the year 2000. He retired as Professor of Electrical Engineering Department of Assam Engineering College. His research area includes Instrumentation and Power System.