ORIGINAL ARTICLE

# Intelligent Fault Diagnosis of Roller Bearings using Bond Graph-Transformer-based Model

**Waqar Ahmad[1], Syed Humayoon Shah[1,*], Naeem ul Islam[2], J L Ordonez[3], Said Ghani Khan[4], Shahbaz Khan[5]**

[1] *College of Electrical and Communication Engineering, Yuan Ze University, Taoyuan City, Taiwan.*

[2] *Department of Computer Science and Engineering, Yuan Ze University, Taoyuan City, Taiwan.*

[3] *Faculty of Engineering, Universidad Tecnológica Centroamericana, San Pedro Sula, Honduras.*

[4] *College of Mechanical Engineering Department, College of Engineering, University of Bahrain, Isa Town, Bahrain.*

[5] *School of Mechanical and Manufacturing Engineering, NUST Islamabad, Islamabad, Pakistan.*

Corresponding author: syedshah@saturn.yzu.edu.tw

(**Received** 17 March 2026, **Revised** 23 March 2026, **Accepted** 07 June 2026)

DOI: 10.5875/m1dcb909

**Abstract:** This study introduces an advanced hybrid framework for intelligent fault diagnosis of bearings by merging Bond Graph-based physical modelling with a Transformer-based deep learning technique. Traditional data-driven methods rely heavily on large, labeled datasets. This limitation is addressed in this study by employing a synthetic vibration signal generated from bond graph-derived differential equations representing various fault conditions, such as healthy, inner race fault, outer race fault, and ball fault. To extract discriminative features from these synthetic signals, a modified Transformer model is created that has been tailored for temporal signal processing. The model was trained and assessed through numerous trials, demonstrating approximately 92.2 % accuracy in both the training and validation phases.

**Keywords:** Bond graph, Deep learning, Transformer.

## 1. Introduction

Rotating machinery is essential in industries like manufacturing, energy, transportation, and aerospace, where faults can cause significant operational losses. Bearings, being critical components, are especially prone to failures, making early and accurate fault diagnosis vital for ensuring reliability and efficiency [1]. Traditional methods, such as model-based [2], signal-based [3], and knowledge-based diagnosis [4], rely on manual feature extraction and expert knowledge, limiting their adaptability. Recently, sensor fusion [5], data-driven, and deep learning approaches [6] have emerged, offering automated feature learning and improved diagnostic accuracy under varying operating conditions.

Deep learning (DL) models rely on large, labeled datasets, but in industrial settings, fault data are often scarce due to the rarity and high cost of inducing faults [7]. To overcome this limitation, synthetic data generation using Bond Graph modeling has gained attention [8]. This physics-based approach models energy interactions in dynamic systems [9], enabling the simulation of realistic fault conditions and the creation of synthetic data that captures the true physical behavior of machinery [10]. Originally developed for natural language processing [11], Transformers excel at modeling sequential data through their self-attention mechanism, making them ideal for time-series vibration analysis in rotating machinery [12]. This study proposes a hybrid approach combining Bond Graph-based synthetic data generation with a Transformer model to overcome data scarcity and improve fault diagnosis accuracy [13]. Earlier fault diagnosis methods relied on signal processing techniques [14] in the time, frequency, and time-frequency domains to extract vibration features. These were then classified using machine learning models like SVMs [15], ANNs [16], and Random Forests [17]. While effective in specific scenarios, these approaches required extensive domain expertise for feature selection [18] and were often sensitive to noise and varying operating conditions.

Deep learning has revolutionized fault diagnosis by enabling automatic feature extraction from raw vibration data. CNNs effectively capture local signal patterns, while RNNs, especially LSTMs, model temporal dependencies

[19], [20], [21]. However, CNNs struggle with long-term dependencies, and RNNs are computationally expensive and prone to vanishing gradients [22]. Moreover, both architectures rely heavily on large, labeled datasets for optimal performance. In various sequence modeling tasks, the transformers have shown dominance in their performance due to their self-attention mechanisms [23]. Analysis of complex time series data [24] can be carried out with ease by the help of transformers due to their capacity to capture global relationships. Recent research has examined the use of Transformers for fault diagnosis [25], [26], [27], [28], and the results show higher accuracy and robustness than typical deep learning models.

To tackle the challenge of limited labeled data, researchers have explored synthetic data generation. GANs can produce realistic fault data to improve model training [29], but their outputs may be complex and lack physical interpretability. Bond Graph modeling offers a physics-based alternative, simulating system dynamics under various fault conditions to generate data that accurately reflects the machine's true physical behavior [30]. Combining physics-based modeling with data-driven learning provides a promising fault diagnosis approach. Integrating Bond Graph-generated data into deep learning models enhances diagnostic performance, especially with small, labeled datasets, by merging the interpretability of physical models with the flexibility of machine learning.

The rest of the article is organized as follows: Section II explains the proposed methodology. Section III discusses the experimental outcomes, and finally, Section IV provides a conclusion and discussion of the research.

## 2. Proposed Methodology

This section outlines the methodology that combines Bond Graph-based physical modeling with a Transformer-based deep learning framework for accurate vibration-based fault diagnosis of rolling element bearings, as illustrated in Figure 1.

To capture the dynamic behavior of the bearing system, a Bond Graph (BG) model was developed to represent energy interactions among mechanical components. The system employed a lumped-parameter approach consisting of inertial (*I*), compliance (*C*), and resistance (*R*) elements. The inertial element models the shaft's rotational mass, the compliance element represents bearing stiffness, and the resistance element accounts for friction and damping. The input torque was modeled as a source of effort ($S_e$), while vibration responses were measured as flow-related outputs. The BG framework facilitates systematic derivation of system dynamics from power-conserving components. In this study, vibration signals were generated for various bearing conditions, such as normal condition, ball fault, inner race fault, and outer race fault across multiple speeds using the Bond Graph-based physical and mathematical model. These vibration signals serve as the foundation for bearing health assessment and fault prediction.

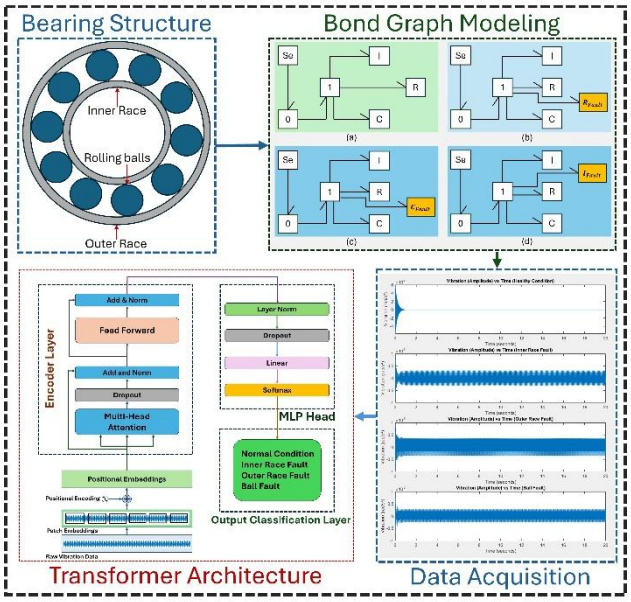


Figure 1. Block Diagram of Proposed Methodology.

### 2.1 Bearing Structure

A rolling element bearing comprises an outer race (stationary, which contacts the housing), an inner race (rotating, which contacts the shaft), rolling elements (balls that reduce friction), and a cage (retainer that spaces and aligns the balls). These components can develop flaws such as surface wear, cracks, or deformation. Figure 2 shows the bearing structure.

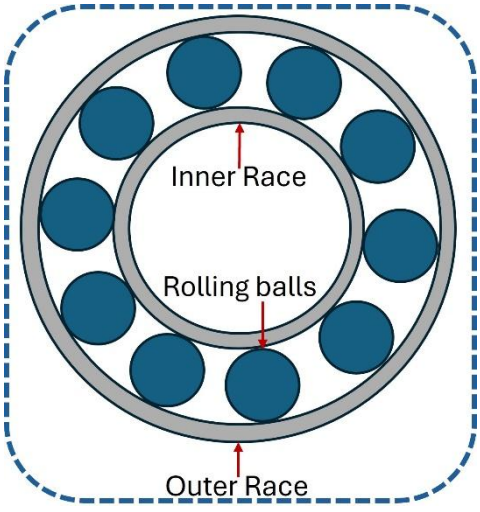


Figure 2. Components of Bearing

Faults produce distinct vibration patterns that can be identified using frequency-domain analysis [31]. Fault frequencies are diagnostic markers for individual bearing faults and are derived using the methods in Equations

1-4. Fault frequencies help pinpoint faulty areas and inform feature extraction.

$$f_{BPFO} = n.f_{rot}.\left(1 - \frac{d}{D}cos\theta\right) \quad (1)$$

$$f_{BPFO} = n.f_{rot}.\left(1 + \frac{d}{D}cos\theta\right) \quad (2)$$

$$f_{BSF} = \frac{d}{2D}.f_{rot}.\left(1 - \left(\frac{d}{D}cos\theta\right)^2\right) \quad (3)$$

$$f_{FTF} = \frac{f_{rot}}{2}.\left(1 - \frac{d}{D}cos\theta\right) \quad (4)$$

$f_{BPFO}$ is ball pass frequency outer race, $f_{BPFI}$ is the ball pass frequency inner race, $f_{BSF}$ is the ball spin frequency, and $f_{FTF}$ is the fundamental train frequency. The variables are ball diameter $d$, pitch diameter $D$, contact angle $\theta$, number of rolling elements $n$, and rotational frequency $f_{rot}$.

### *2.2 Bond Graph Modelling and Simulation*

This section focuses on the Bond Graph modelling and simulation approach utilized in our research for fault diagnostics of rolling element bearings. To simulate the dynamic behaviour of rolling element bearings in various health states, we used Bond Graph theory, a sophisticated domain-independent graphical representat -ion of physical dynamic systems. This method enables us to capture energy interactions between mechanical components and investigate how faults affect system dynamics. Bond graphs employ standardized elements to represent power exchange within a system. The fundamental elements of a mechanical bearing system include $S_e, 0_{junction}, 1_{junction}, I, R, C$. $S_e$ is a source of effort (external force input), $0_{junction}$ express common effort nodes (parallel connection), $1_{junction}$ is used for common flow nodes (series connection), $I$ is used as an inertial element and illustrates the mass or moment of inertia (rotating element, balls), *R* is a resistive element which models the damping or friction (internal resistance of the bearing), and *C* is a capacitive element utilized for compliance or elasticity (deformation of the raceways). The Bond graph (BG) for a bearing is showcased in Figure 3, where we have four categories: a, b, c, and d.

Although Equations (1) - (4) are derived from bearing kinematics to determine characteristic defect frequencies, these frequencies are incorporated into the Bond Graph framework as periodic excitation sources associated with localized fault-induced dynamic disturbances. The Bond Graph model itself represents the physical energy interactions of the bearing system through inertia, damping, and stiffness elements, while the kinematic fault frequencies characterize the periodic excitation generated by defect interactions during

Figure 3. Bearing Bond graph for different Faults

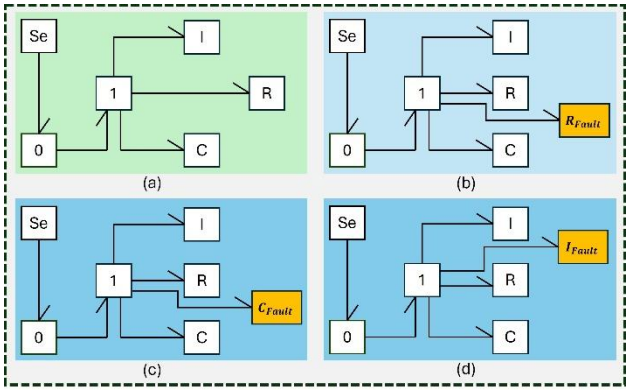


Figure 3. Bearing Bond graph for different Faults

operation. Therefore, the proposed framework combines kinematic fault characterization with energy-based dynamic modeling to generate physically meaningful vibration responses for fault diagnosis

Category 'a' presents BG for bearing in a healthy condition. Part 'b' illustrates the inner race fault, 'c' describes the outer race fault, and 'd' shows the ball fault. A popular method in dynamic mechanical system analysis, the mass-spring-damper analog system, is used in this study to model the bearing system. Newton's Second Law of Motion is expressed in Equation 5, which governs the dynamic behavior of the system.

$$M^*\ddot{x}(t) + R^*\dot{x}(t) + \frac{1}{C^*}x(t) = F \quad (5)$$

Where *M*, R*, and C** denote the effective mass, damping, and compliance parameters that fluctuate according to the fault condition, *t* is the time duration, and *F* is the force. According to these conditions, when we solve the equations, we will get separate equations for different faults. For normal condition, Equation 5 will remain the same. After solving Equation 5 for the inner race fault, we get Equation 6. Damping *R* will increase for the inner race fault due to friction or damage to the inner race fault due to friction or damage to the inner race.

$$M^*\ddot{x}(t) + (R + \Delta R)\dot{x}(t) + \frac{1}{C^*}x(t) = F \quad (6)$$

Stiffness will reduce and compliance *C* will increase due to looseness, and the reason for this is spalling at the outer race, resulting in outer race fault as shown in Equation 7.

$$M^*\ddot{x}(t) + R^*\dot{x}(t) + \frac{1}{(C+\Delta C)}x(t) = F \quad (7)$$

When a ball fault occurs, the mass and inertia will increase due to damage to the ball, which results in Equation 8.

$$(M + \Delta M)\ddot{x}(t) + R^*\dot{x}(t) + \frac{1}{C^*}x(t) = F \quad (8)$$

Mathematical modeling is carried out, and Equations 5 to 8 were extracted. These equations express the exact same state of the bearing during different working conditions. To validate the proposed BG-generated synthetic dataset, comparative analyses between experimentally measured bearing vibration signals in CWRU (Case Western Reserve University) bearing dataset and BG-generated signals were conducted using the 12 kHz sampling rate in both the time and frequency domains. The comparison demonstrated that the generated signals preserve representative oscillatory behavior, impulsive fault characteristics, and dominant spectral components corresponding to different bearing fault conditions, thereby confirming the suitability of the proposed BG framework for Transformer-based fault diagnosis.

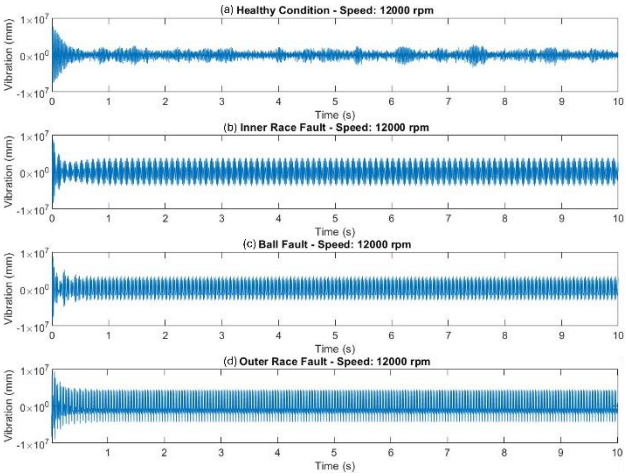


Figure 4. Time-domain vibration signals generated using the proposed Bond Graph (BG)-based dynamic model under (a) healthy (b) inner-race fault (c) ball fault (d) outer-race fault conditions at 12000 rpm.

Figures 4-7 present the comparative time-domain and frequency-domain analyses of both the Bond Graph (BG)-generated synthetic vibration signals and the experimentally measured CWRU bearing signals under healthy, inner-race fault, ball fault, and outer-race, ball fault, and outer-race fault conditions at 12000 rpm.

Figures 4 and 5 illustrate the time-domain responses of the BG-generated and real bearing vibration signals, respectively. The generated signals exhibit oscillatory and impulsive characteristics similar to those observed in the experimental measurements, particularly under faulty operating conditions where periodic fault-induced impacts become more prominent. Similarly, Figures 6 and 7 present the corresponding frequency-domain power spectrum analyses. The BG-generated signals successfully preserve the dominant spectral peaks and harmonic fault related frequency components observed in the real bearing data. The healthy condition demonstrates relatively stable low energy spectral behavior, whereas faulty conditions produce distinct high energy spectral components associated with localized fault excitation frequencies. These observations confirm that the proposed BG framework can effectively reproduce representative dynamic fault characteristics and generate physically meaningful synthetic vibration data suitable for Transformer-based intelligent fault diagnosis.

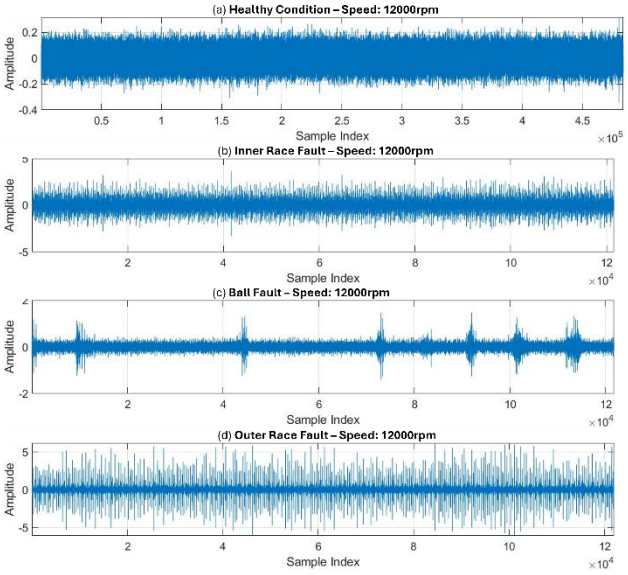


Figure 5. Time-domain vibration signals obtained from the CWRU experimental bearing dataset under (a) healthy (b) inner-race fault (c) ball fault (d) outer-race fault conditions at 12000 rpm.

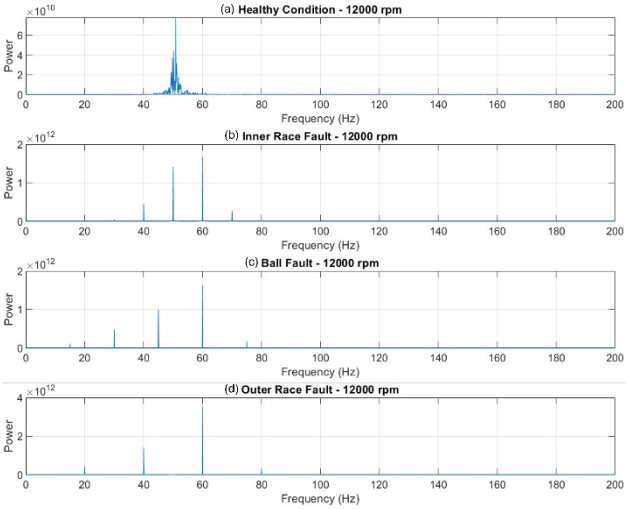


Figure 6. Frequency-domain power spectrum analysis of vibration signals generated using the proposed Bond Graph (BG)-based dynamic model under (a) healthy (b) inner-race fault (c) ball fault (d) outer-race fault conditions at 12000 rpm.

## *2.3 Data Acquisition and Pre-Processing*

Conventional condition monitoring relies on vibration sensors mounted on rotating machines. In this study, a Bond Graph-based physics simulation is used to generate synthetic vibration signals, allowing controlled fault experiments without noise, cost, or time of real-world measurements. Bond Graphs offer a systematic and efficient way to model multi-domain

physical systems based on energy conservation. In this study, the bearing is modeled as a second-order linear dynamic system with mass (*M*), damping (*R*), and stiffness (*C*). The Bond Graph produces the governing equation derived from Newton's second law of motion as presented in Equation 5. Bearing faults are simulated by varying *M*, *R*, and *C* to represent different degradation levels and fault severities.

This physics-based fault injection approach produces clear and distinct dynamic responses for different fault conditions. Each model equation is numerically solved in MATLAB using the *ODE45* solver over a *t=10 seconds* time period with a *12 kHz* sampling rate, generating time-domain displacement (vibration amplitude) $x(t)$ and velocity $\dot{x}(t)$. To ensure fair comparison, all simulations share identical initial conditions and external force input $x(t)$. Data were generated at 4000, 8000, and 12000 rpm, allowing the model to learn from variable speeds for robust fault diagnosis. The resulting dataset includes healthy, inner race fault, outer race fault, and ball fault signals at each speed, all saved as *.mat* files.

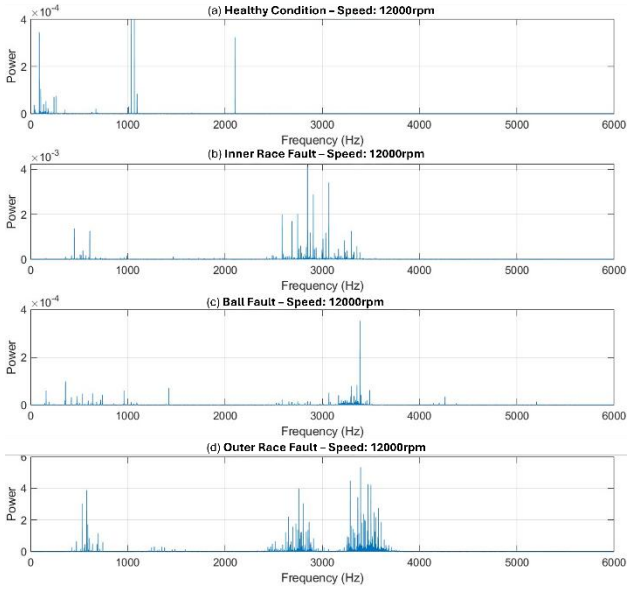


Figure 7. Frequency-domain power spectrum analysis of experimentally measured bearing vibration signals from the CWRU dataset under (a) healthy (b) inner-race fault (c) ball fault (d) outer-race fault conditions at 12000 rpm.

### *2.4 Deep Learning Transformer-based Architecture*

In this research, we have employed a modified Transformer-based model, DSC Transformer, such that it utilizes vibration data acquired from the bond graph of bearings and employs this data for fault diagnosis. This architecture is inspired by the original Transformer encoder design; however, the suggested model is altered for the classification of time series vibration data so that it employs the combination of patch embedding, positional encoding, and a stacked encoder design.

Unlike recent Transformer-based meta-learning approaches designed for few-shot adaptation under limited data conditions, the proposed framework integrates Bond Graph (BG)-based physical system modeling with Transformer-based deep feature learning for bearing fault diagnosis. The proposed method does not employ episodic meta-learning or MAML-based optimization; instead, physically interpretable vibration responses generated through BG dynamic modeling are directly utilized as inputs to the Transformer encoder. This hybrid physics-informed framework enables the Transformer model to capture fault-related temporal dependencies while preserving the physical significance of stiffness, damping, and excitation interactions within the bearing system, thereby differentiating the proposed approach from purely data-driven Transformer meta-learning methods.

The Transformer architecture used in this study is shown in Figure 8. The model comprises four encoder layers and is trained on data from 10 operating conditions: one healthy and three fault types (inner race, outer race, and ball fault) at 4000, 8000, and 12000 RPM. Each vibration signal is segmented into 1024 samples, which are further divided into 16 non-overlapping patches.

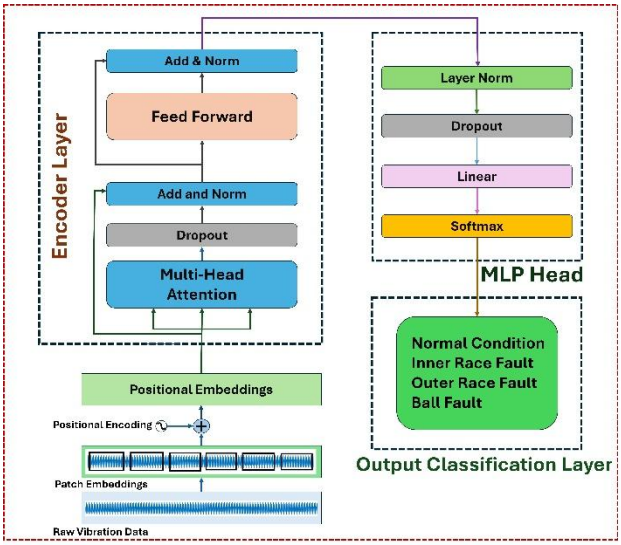


Figure 8. Transformer architecture for bearing fault diagnosis

A linear projection converts each patch into a fixed-dimensional embedding vector of size 64, capturing temporal features such as amplitude modulation, frequency content, transient behaviour, and statistical properties. The input signal with *L = 1024* is divided into *P=16* patches, where $x = [x_1, x_2, \ldots\ldots x_{16}]$, and each patch $x \in \mathbb{R}^{\frac{L}{P}}$ is mapped to each embedding through a learnable linear projection.

$$z_i = [W_e x_i + b_e, \quad z_i \in \mathbb{R}^d] \tag{9}$$

Where $x_i$ denote the $i$-th input patch vector extracted from the original time series. It has dimension $\frac{L}{P}$, where

*L = 1024* is the total length of the signal and *P = 16* is the number of patches, resulting in $x \in \mathbb{R}^{64}$. Each patch is linearly projected into a higher-dimensional latent space using a learnable weight matrix $W_e \in \mathbb{R}^{d \times \frac{L}{P}} = \mathbb{R}^{64 \times 64}$, along with a learnable bias vector $b_e \in \mathbb{R}^d = \mathbb{R}^{64}$. The resulting embedding vector for each patch is computed in Equation 9, where $z_i \in \mathbb{R}^d = \mathbb{R}^{64}$. Here, *d = 64* represents the model dimension, i.e., the size of the embedding space. Information about the sequence of the input embeddings is provided to preserve the order of the input embeddings. However, a specific class token is added to the patch sequence by employing positional encoding to keep the order of the sequences. $Z$ is the input embedding, where *PE* is the positional encoding matrix as expressed in Equation 10.

$$Z = [z_{cls}, z_{1,} z_{2,} z_{3,} \dots \dots z_{16}] \quad (10)$$

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (11)$$

The transformer model utilized in this research employs four encoder layers *N=4*. Each layer includes a multi-head self-attention and a Feedforward network. Self-attention block uses the mathematical expression in Equation 11 to extract the value of *Q*(Queries), *K*(Keys), and *V*(Values). Each sub-block is preceded by a residual connection and layer normalization expressed in Equations 12 and 13.

$$Z' = LayerNorm(Z + MultiHead(Z)) \quad (12)$$

$$Z'' = LayerNorm(Z' + FFN(Z')) \quad (13)$$

Where $Z'$ is multi-head self-attention and $Z''$ is a feed-forward network. Four heads are used by each attention layer, which allows the model to attend to multiple representation subspaces at the same time. The model dimension is projected from *d=64* to *d=256* with the help of the feedforward layers, as showcased in Equation 14.

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (14)$$

$$y = Softmax(W_c z_{CLS} + b_c) \quad (15)$$

To prevent overfitting and improve generalization, a dropout is applied with a rate of 0.1. The representation related to the class tokens is extracted after passing through all layers of encoder and is sent to the classification head where the final MLP layer produces the class logits for various health conditions of the bearing. Where *Y* is the output prediction vector typically class probabilities used for classification. Softmax is the activation function that converts logits into a probability distribution over classes. $W_c$ is the weight matrix of the final output classification layer. $Z_{CLS}$ is the embedding of the special class token output by the encoder and $b_c$ is Bias vector of the final classification layer.

## 3. Results

This section evaluates the proposed Bond-Graph–Transformer fault diagnosis model. The model was trained for 100 epochs, with training accuracy steadily rising to near 100% and training loss approaching zero, indicating accuracy reached 99% after 70 epochs, with a consistent decline in validation loss, showing strong generalization without overfitting as illustrated in Figure 9. The Transformer has four encoder layers, a model dimension of 64, and a sequence length of 1024, with input signals split into 16 patches to capture temporal dependencies. Training on an NVIDIA RTX 3060 GPU took about 5 seconds per epoch (9 minutes total) with a batch size of 16 and 4 effective feature learning and optimization. Validation GB GPU memory. Despite its lightweight design, the model achieved 92.2% classification accuracy, demonstrating high performance, efficiency, and suitability for real-time industrial applications. Detailed results are expressed in Table I.

Table 1. Results for the proposed methodology

| Model | Number of Epochs | Accuracy |
|---|---|---|
| Transformer | 100 | 92.2% |

The separability of the features learned from a high-dimensional feature representation were assessed with the help of a t-SNE (t-Distributed Stochastic Neighbor Embedding) projection. The results are showcased in Figure 10. which demonstrates the distinct and well-separated clusters for each class label. The noticeable separation of these clusters in 2D space signifies that the Transformer-based model has successfully acquired class-specific embeddings, allowing the final classification process much easier for the output MLP layer. The overall accuracy of the model is 92.2 %. The effectiveness of the proposed BG modelling for data acquisition combining with Transformer-based DL architecture for fault diagnosis is proved from the results. The model has displayed promising results for real-time fault diagnosis with high accuracy in rotating machinery. The confusion matrix showcased in Figure 11 demonstrates the classification results for a multi-class bearing fault diagnosis system having an overall accuracy of 92.2%.

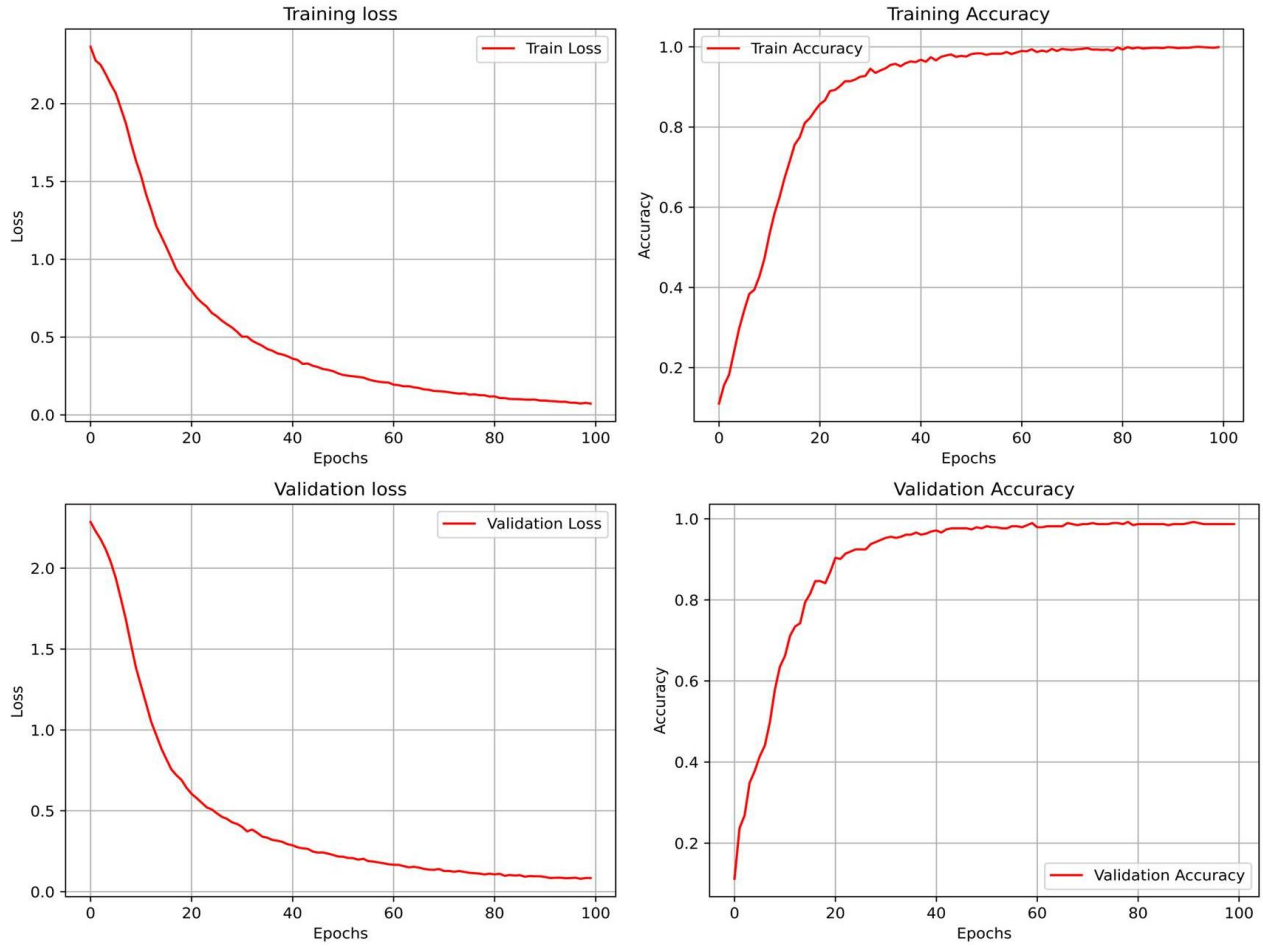


Figure 9. Result of the BG-TF-based fault diagnosis of bearings.

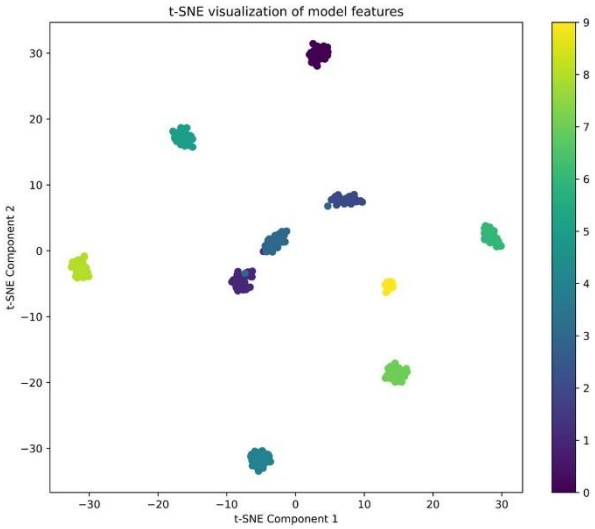


Figure 10. t-SNE plot for feature separability

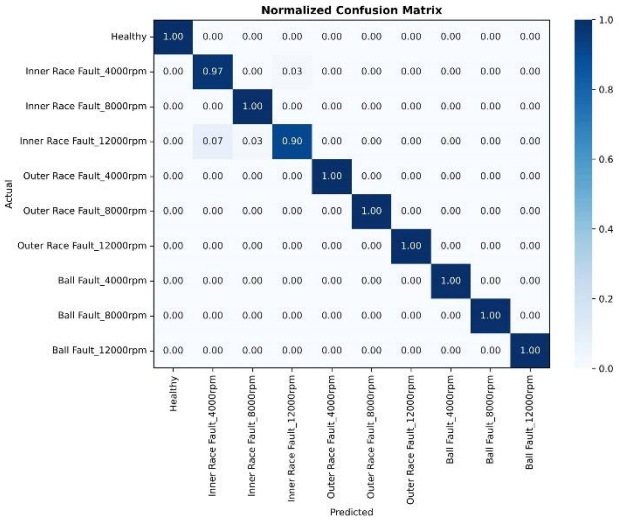


Figure 11. Confusion matrix for the accuracy of the Transformer model.

The matrix contains ten fault categories such as Healthy, Inner Race Faults, Outer Race Faults, and Ball Faults, each with three rotating speeds (4000, 8000, and 12000 rpm). The matrix displays high diagonal

dominance, clearly indicating that the model correctly predicted the majority of data in their respective classes.

## 4. Conclusion and Discussion

This study presents a hybrid fault-diagnosis approach that integrates Bond Graph (BG) modeling with a deep-learning Transformer architecture to detect and classify bearing faults from vibration data. The proposed method addresses two major challenges in predictive maintenance: the scarcity of labeled fault data and the need for accurate real-time diagnosis. The data scarcity issue is mitigated through Bond Graph-based simulation of diverse fault conditions, while the Transformer model enables precise and efficient fault classification. Model generalization and performance were validated using various analysis, including accuracy metrics, loss evaluation, and t-SNE visualizations. The suggested framework is ideal for real-world industrial applications such as condition-based maintenance. Multi-sensor fusion, transfer learning for multiple rotating machinery faults, online learning, and experimental validation with real vibration data could be included in the future to extend the research [32].

Although the present study primarily focuses on the development and validation of the proposed BG-Transformer framework, comprehensive comparative evaluations against additional state-of-the-art fault diagnosis methods remain an important direction for future research. Future work will investigate detailed benchmarking studies involving conventional machine learning and deep learning approaches under identical operating conditions to further evaluate the generalization capability and diagnostic robustness of the proposed framework.

## 5. Acknowledgment

We acknowledge the use of AI-assisted tools to improve the readability, clarity, and formatting of this research manuscript. These methods were used exclusively for language improvement and did not influence the scientific content, analysis, or interpretation of the research results.

Declaration: None

Funding Statement:

This study was financially supported by the Ministry of Science and Technology, Taiwan (R.O.C), under grant number 114-2222-E-155-001.

Conflict of Interest:

There is no conflict of interest between the authors.

Authors' Contributions:

All authors have made equal contributions to this article.

Author Disclosure Statement:

The authors declare that they have no competing Interests.

## References

[1] S. Karabay and I. Uzman, "Importance of early detection of maintenance problems in rotating machines in management of plants: Case studies from wire and tyre plants," *Eng. Fail. Anal.*, vol. 16, no. 1, pp. 212–224, 2009, doi: 10.1016/j.engfailanal.2008.03.003.

[2] R. Isermann, "Model-based fault-detection and diagnosis - Status and applications," *Annu. Rev. Control*, vol. 29, no. 1, pp. 71–85, 2005, doi: 10.1016/j.arcontrol.2004.12.002.

[3] P. Gangsar and R. Tiwari, "Signal-based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review," *Mech. Syst. Signal Process.*, vol. 144, p. 106908, 2020, doi: 10.1016/j.ymssp.2020.106908.

[4] Y. Chi, Y. Dong, Z. J. Wang, F. R. Yu, and V. C. M. Leung, "Knowledge-Based Fault Diagnosis in Industrial Internet of Things: A Survey," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 12886–12900, 2022, doi: 10.1109/JIOT.2022.3163606.

[5] F. Kibrete, D. Engida Woldemichael, and H. Shimels Gebremedhen, "Multi-Sensor data fusion in intelligent fault diagnosis of rotating machines: A comprehensive review," *Meas. J. Int. Meas. Confed.*, vol. 232, no. April, p. 114658, 2024, doi: 10.1016/j.measurement.2024.114658.

[6] S. Qiu *et al.*, "Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review," *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031305.

[7] M. Kahr, G. Kovács, M. Loinig, and H. Brückl, "Condition Monitoring of Ball Bearings Based on Machine Learning with Synthetically Generated Data," *Sensors*, vol. 22, no. 7, pp. 1–17, 2022, doi: 10.3390/s22072490.

[8] M. Taktak, S. Triki, and A. Kamoun, "Real-time algorithm based on time series data abstraction and hybrid bond graph model for diagnosis of switched system," *Eng. Appl. Artif. Intell.*, vol. 59, no. December 2016, pp. 51–72, 2017, doi: 10.1016/j.engappai.2016.12.009.

[9] E. G. Ovy and Q. Sun, "Ball-bearing dynamics modelling by a bond graph approach," *Int. J. Model. Simul.*, vol. 42, no. 1, pp. 17–33, 2022, doi: 10.1080/02286203.2020.1818426.

[10] M. Nakhaeinejad and M. D. Bryant, "Dynamic modeling of rolling element bearings with surface contact defects using bond graphs," *J. Tribol.*, vol. 133, no. 1, pp. 1–12, 2011, doi: 10.1115/1.4003088.

[11] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[12] Q. Snyder, Q. Jiang, and E. Tripp, "Integrating

self-attention mechanisms in deep learning: A novel dual-head ensemble transformer with its application to bearing fault diagnosis," *Signal Processing*, vol. 227, no. April 2024, p. 109683, 2025, doi: 10.1016/j.sigpro.2024.109683.

[13] C. Y. Tai and Y. Altintas, "A hybrid physics and data-driven model for spindle fault detection," *CIRP Ann.*, vol. 72, no. 1, pp. 297–300, 2023, doi: 10.1016/j.cirp.2023.04.054.

[14] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribol. Int.*, vol. 96, pp. 289–306, 2016, doi: 10.1016/j.triboint.2015.12.037.

[15] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018, doi: 10.1016/j.neucom.2018.05.002.

[16] A. A. A. Mohd Amiruddin, H. Zabiri, S. A. A. Taqvi, and L. D. Tufa, "Neural network applications in fault diagnosis and detection: an overview of implementations in engineering-related systems," *Neural Comput. Appl.*, vol. 32, no. 2, pp. 447–472, 2020, doi: 10.1007/s00521-018-3911-5.

[17] K. Guo, X. Wan, L. Liu, Z. Gao, and M. Yang, "Fault diagnosis of intelligent production line based on digital twin and improved random forest," *Appl. Sci.*, vol. 11, no. 16, 2021, doi: 10.3390/app11167733.

[18] S. Buchaiah and P. Shakya, "Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection," *Meas. J. Int. Meas. Confed.*, vol. 188, no. August 2021, 2022, doi: 10.1016/j.measurement.2021.110506.

[19] D. Ruan, J. Wang, J. Yan, and C. Gühmann, "CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis," *Adv. Eng. Informatics*, vol. 55, no. October 2022, p. 101877, 2023, doi: 10.1016/j.aei.2023.101877.

[20] C. Eang and S. Lee, "Predictive Maintenance and Fault Detection for Motor Drive Control Systems in Industrial Robots Using CNN-RNN-Based Observers," *Sensors*, vol. 25, no. 1, 2025, doi: 10.3390/s25010025.

[21] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," *J. Intell. Manuf.*, vol. 32, no. 4, pp. 971–987, 2021, doi: 10.1007/s10845-020-01600-2.

[22] D. J. Choi, J. H. Han, S. U. Park, and S. K. Hong, "Comparative Study of CNN and RNN for Motor fault Diagnosis Using Deep Learning," *2020 IEEE 7th Int. Conf. Ind. Eng. Appl. ICIEA 2020*, pp. 693–696, 2020, doi: 10.1109/ICIEA49774.2020.9102072.

[23] J.-K. Kämäräinen, "Introduction to Sequence Modeling with Transformers," 2025, [Online]. Available: http://arxiv.org/abs/2502.19597

[24] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, R. P. Ramachandran, and G. Rasool, "Transformers in Time-Series Analysis: A Tutorial," *Circuits, Syst. Signal Process.*, vol. 42, no. 12, pp. 7433–7466, 2023, doi: 10.1007/s00034-023-02454-8.

[25] Y. Hou, J. Wang, Z. Chen, J. Ma, and T. Li, "Diagnosisformer: An efficient rolling bearing fault diagnosis method based on improved Transformer," *Eng. Appl. Artif. Intell.*, vol. 124, no. May, p. 106507, 2023, doi: 10.1016/j.engappai.2023.106507.

[26] A. Y. Zhou and A. Barati Farimani, "FaultFormer: Pretraining Transformers for Adaptable Bearing Fault Classification," *IEEE Access*, vol. 12, no. April, pp. 70719–70728, 2024, doi: 10.1109/ACCESS.2024.3399670.

[27] S. Zhang, J. Zhou, X. Ma, S. Pirttikangas, and C. Yang, "TSViT: A Time Series Vision Transformer for Fault Diagnosis of Rotating Machinery," *Appl. Sci.*, vol. 14, no. 23, 2024, doi: 10.3390/app142310781.

[28] J. Z. X. M. S. P. C. Y. Shouhua Zhang, "TSViT A Time Series Vision Transformer for Fault Diagnosis," 2023.

[29] B. Vaz and Á. Figueira, "GANs in the Panorama of Synthetic Data Generation Methods," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 21, no. 1, 2024, doi: 10.1145/3657294.

[30] W. Borutzky, "A hybrid bond graph model-based - Data driven method for failure prognostic," *Procedia Manuf.*, vol. 42, pp. 188–196, 2020, doi: 10.1016/j.promfg.2020.02.069.

[31] S. A. McInerny and Y. Dai, "Basic vibration signal processing for bearing fault detection," *IEEE Trans. Educ.*, vol. 46, no. 1, pp. 149–156, 2003, doi: 10.1109/TE.2002.808234.

[32] R. Isermann, "On fuzzy logic applications for automatic control, supervision, and fault diagnosis," *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans.*, vol. 28, no. 2, pp. 221–235, 1998, doi: 10.1109/3468.661149.