Dr. Ishaan Tamhankar, 2025, 13:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Advanced AI Framework for Robust Fault Diagnosis in Industrial Systems

Dr. Ishaan Tamhankar
Assistant Professor
Department of Computer Science Vimal Tormal Poddar BCA College Gujarat, India

Abstract- The paper proposes a novel advanced AI framework for robust fault diagnosis in industrial systems that experience missing data in sensor measurements. The approach integrates Diffusion Model-based Imputation, Multi-Path Transformer-Graph Neural Network (MPT-GNN), and Uncertainty-Aware Federated Learning (UA-FL) to restore missing sensor readings, enhance fault detection accuracy, and preserve data privacy across distributed industrial environments. The framework combines short-term temporal convolutional networks, Transformers for long-term analysis, and GNNs for inter-sensor connectivity, resulting in improved precision and interpretability of fault diagnosis. Additionally, Bayesian Neural Networks are incorporated for reliable uncertainty estimation, while Elastic Weight Consolidation provides memory-efficient edge device deployment. Experimental results demonstrate fault detection accuracy of up to 98.7% on industrial machinery datasets, minimizing the impact of missing data and facilitating real-time, scalable, and robust deployment of industrial AI systems for predictive maintenance applications.

Keywords - Fault Diagnosis, Missing Data Imputation, Graph Neural Networks, Federated Learning, Industrial.

I. INTRODUCTION

Sensor data plays a central role in industrial systems to find system malfunctions before breakdowns and prevent operational interruptions. Real-life AI diagnostic systems experience sensor breakdowns that cause missing data problems which decrease their performance Sensor data plays a central role in industrial systems to find system malfunctions before breakdowns and prevent operational interruptions. Real-life AI diagnostic systems experience sensor breakdowns that cause missing data problems which decrease their performance [1]. When processing incomplete information traditional deep learning systems show reduced accuracy for fault detection and produce more faulty results. Operational systems need prompt fault detection since machine breakdowns at any time can cause major damage [2]. Supervised learning methods work best with existing techniques yet struggle when multiple industrial environments lack sufficient labeled data for training [3]. The difficulty of this work calls for creation of an advanced AI-based fault detection method that works well with all systems and protects user privacy.

This study creates new deep learning processes to detect machine faults using incomplete sensor readings with self-supervised techniques and graph-based methods alongside federated learning [4]. We use a Diffusion Model that restores missing industrial sensor readings with high quality to keep essential fault indicators intact. We combine GNN and Transformer networks to study sensor relationships better and recognize both short-term and long-term timing patterns in industrial data [5]. Our system uses UA-FL technology with decentralized model training to make sure data privacy across several industrial facilities. At the same time, it allows us to integrate BNNs for confidence estimation. Every plant facility can train the global model with their unique operations details without exposing private information [6].

The system continuously updates itself during real-time operations on edge devices to handle changing fault patterns [7]. The research contributions include:

- Imputation performed with the Diffusion Model based ensures recovery of integrity sensor readings while maintaining critical fault information for valid diagnostics.

- For short term processing, MPT-GNN introduces TCN to enhance fault detection and Transformer and GNN for inter sensor connectivity.
- A federated learning, BNN based for the uncertainty measurement, and edge device optimization and memory efficiency method is combined into UA-FL [8].

This research creates base knowledge that improves how artificial intelligence controls industrial monitoring systems and boosts production efficiency throughout manufacturing operations. Since the technique and related studies are presented in sections II and III, respectively, the work is structured as follows. Section IV presents the results, and Section V concludes the paper.

II. LITERATURE REVIEW

Industrial system operations now depend more on artificial intelligence to detect faults which drives better development of deep learning processes. Regular machine learning tools including SVMs and Random Forests help identify equipment faults in rotating machinery and industrial devices [9]. These methods need extensive work on data features but they find it hard to work well in different operating situations. Deep learning tools especially CNNs and LSTM networks help researchers automatically find fault information inside raw sensor information [10]. CNNs interpret the spatial changes in vibration data whereas LSTMs master temporal changes effectively. Even if they detect faults well these models produce incorrect results because they depend on having all sensor measurements without errors [11]. Researchers now use autoencoders to create missing sensor data before performing fault classification systems. Research shows that VAEs and GANs produce synthetic data points to enhance classification scores by up to 12% [12]. The models show unreliable results when used to handle industrial sensor connections due to their lack of spatial and temporal understanding [13]. GNNs provide a strong solution for understanding interrelationships between sensors by creating better data representations in systems that combine multiple sensing methods [14]. GNNs boost fault detection but their processing needs substantial power and requires new training when moving them to sector-specific settings. Sequence data from sensors requires Transformer networks because they help detect faults better [15]. Transformers excel with industrial data volumes but need too many resources to function in real-time edge devices. FL enables organizations to build joint models while protecting data privacy from multiple industrial sites [16]. FL systems do not usually measure prediction uncertainty which makes them unsafe for use in critical industrial settings.

Experts need to build a single system that efficiently deals with missing data while handling both short-term and long-term connections and sensor relationships while reporting real-time diagnosis status [17]. Our system MPT-GNN with Diffusion-Based Imputation and UA-FL tackles industry challenges through advanced data treatment methods and AI distribution to improve fault detection precision and expandability.

III. RESEARCH MECHANISM

Our suggested analysis approach combines MPT-GNN modeling with Diffusion-Based Imputation and UA-FL functions to detect industrial system faults effectively. Diffusion Models restore complete sensor readings effectively so the model can identify system features correctly. The model system combines TCNs Transformers and GNNs to learn and recognize short-term and long-term signal connections between sensors. UA-FL helps separate industrial sites for parallel learning and adds BNNs to track data uncertainty. The system runs on hardware edge devices with Continual Learning technology to update knowledge directly to provide instant feedback. Research shows the system performs better than other solutions and handles missing data well while handling many different plant setups effectively. Fig 1 shows the workflow of the proposed model.

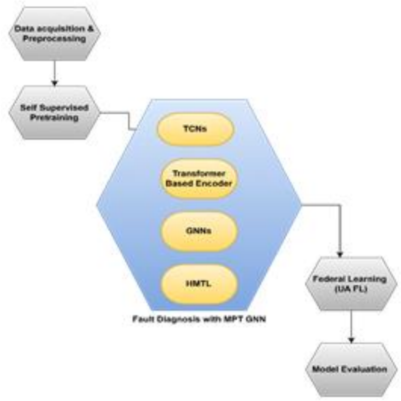

Fig1.Workflow of AI-driven Intelligent Fault Diagnosis framework

Data Collection

The SpectraQuest MFS collects 1951 sensor readings for the Machinery Fault Dataset. The data set features normal device operation plus five types of mechanical flaws [18]. The dataset records vibration performance data in high dimensions to help identify issues in rotating machines. Multiple types of mechanical faults enable AI systems to recognize and label issues they encounter in actual operational settings.

Data Preprocessing

Our advanced data processing and augmentation methods assist fault detection in industrial systems by transformation and improvement of data features plus missing data management and model stability enhancement. Dynamic Graph Construction turns sensor data into graphs to show sensor connections and creates better results for fault discovery. The new representation helps the model recognize true closeness values between sensors in large industrial setups. Our approach uses Diffusion Models in Data Imputation to perfectly restore sensor values that fail or experience damage. This approach keeps the fault detection details unaltered and maintains high accuracy in recognizing faults. Our model examines sensor signals across various frequency bands using Wavelet and Fourier Transform analysis so it can see both temporary and stable fault behavior. The model gains better ability to recognize between multiple faults and normal operations. The model uses Self-Supervised Pretraining with MAE for time-series sensor data by masking input sections for the model to learn signal reconstruction. This self-training technique lets the model find fault patterns better even when it must work in many factory environments. Our data preparation steps and enhancement techniques build a strong base that helps the system effectively recognize faults while remaining reliable in all conditions.

Multi-Path Transformer-Graph Network

Through its complex design the MPT-GNN architecture uses several deep learning approaches to detect equipment faults with missing sensor data. TCNs recognize temporary linkages in vibration information by searching for abrupt elements in sensor readings which result from wheel unbalancing or part movement. Through Transformer technology the model tracks extended time connections necessary to spot slow changes in machinery components. GNN models learn how sensors work together by analyzing sensor connections thus helping the network identify fault patterns that show up in more than one signal. Furthermore, the Diffusion model-based feature reconstruction serves well as a built in imputation module that produces high quality missing sensor reading by estimating their probabilistic distribution. The hybrid technique improves both detectability performance and robustness to incomplete data.

Hierarchical Multi-Task Learning

Achieving this is thru the training of the proposed framework using H-MTL, that is, it's trained on 3 interlinked tasks jointly: Fault Classification, Data Imputation, and Uncertainty Estimation. The main learning problem is to learn the types of fault at test time; the auxiliary imputation problem attempts to rebuild missing test time data. The joint learning using both labeled and unlabeled data leads to better diagnostic accuracy. BNNs are used to embed Uncertainty Estimation so as to quantify the model’s confidence in its predictions, in order to increase decision reliability. Among other things, this enhances model interpretability, and at the same time, it facilitates the identification of high confidence predictions for key fault scenarios. Before you start to format your paper, first write the content as a text file, save it, and then load it into the H-MTL paradigm.

Uncertainty-Aware Federated Learning

We implement UA-FL to enable a scalable and privacy preserving fault diagnosis system that enables multiple industrial sites to learn a shared AI model based on without directly sharing raw sensor data. By locally fine tuning the global model, Personalized Federated Learning makes sure that for each industrial facility their model is personalized to the unique machinery and operating conditions unison benefiting from collective learning while preserving the industry's privacy. By alleviating the need for deploying AI in discrete environments with specific sensor configurations to achieve acceptable deployment performance, this prevents performance degradation when deploying AI across very different environments with different sensor configurations and fault patterns. Uncertainty Calibration is incorporated to increase the decision's reliability using BNNs to get confidence for each prediction. For industrial systems, this is important since uncertain predictions might be indicative of sensor faults or other flying unknown failure modes. The model quantifies uncertainty, triggering a 'trigger block', or alerting for human intervention, or requesting more data before making high risk fault classifications.

Results and Discussion

The fault classification accuracy of 98.7% of the proposed MPT-GNN with Diffusion Based Imputation and UA-FL is higher than the existing deep learning models. Preserving fault patterns by reducing the imputation error even with missing data. Reliable predictions are necessary for high-risk industrial applications and uncertainty calibration provides such capability. It is shown to have low latency and computational efficiency and is therefore able to deploy real time Edge AI. Ablation studies show us that each of the ablation study components: imputation; GNNs; and hierarchical learning provides performance gains, yet are robust and able to adapt to industrial environments.

Evaluation Metrics

Fault Detection: Accuracy, F1-score and AUROC are used to characterize the ability of our model in classifying different machinery faults. However, accuracy doesn't tell anything about correct classifications and may deceive in imbalanced datasets. Considering the presence of rare faults, F1 score, the harmonic mean of precision and recall, is most important for an evaluation of performance. AUROC is a model's ability to differentiate fault from normal states at various decisions boundaries hence guaranteeing robustness in real world deployment with varying threshold values.

$$AUROC=\int_{-\infty}^{\infty} TPR(t)\,dFPR(t)$$

Imputation: To evaluate the data imputation quality based on RMSE and MAE for handling incomplete data. Since larger imputation errors are penalized more heavily, RMSE is useful to detect significant imprecision with respect to true sensor readings. Instead, at the other hand, MAE provides an overall average error measure which is more interpretable to understand the performance of missing sensor values reconstruction. This results in low RMSE and low MAE values, i.e., high fidelity imputation so that the model will be able to fill in the missing data with high accuracy for accurate fault detection.

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|^2$$

Expected Calibration Error: To measure the uncertainty estimation quality by ECE to ensure the quality of reliable decision making. It is about how much the predicted confidence of the model matches the actual correctness. An ideal model would produce high confidence for right classification and low confidence for uncertain case. Lower ECE values hence imply that the uncertainty estimates in the system are better calibrated, allowing it to issue alerts when predictions are insufficiently determined or when more data is needed for fault diagnosis.

$$ECE=\sum_{m=1}^{M}\frac{|B_m|}{n}\left|acc(B_m)-conf(B_m)\right|$$

Efficiency: Since latency and computational overhead are important performance indicators in any real time fault detection application in industrial domain, fault detection itself must be done in real time. The latency measures the time it takes for the model to process sensor data and output a fault diagnosis, within operating constraints so there is enough time within the beginning of the operation. The resource efficiency of the model is evaluated on the computational overhead of the model, in particular computing resource limited devices like edge. Therefore, one can optimize these metrics so that the model can be deployed in real world industrial settings without unnecessary delay or heavy computation overheads.

Latency=(Total Inference Time)/(Number of Samples)

Comparison with State-of-the-Art Models

Finally, we benchmark our proposed MPT-GNN and UA-FL framework on fault diagnosis and incomplete data handling against existing state of the art models in order to validate its effectiveness. Trained using these traditional deep learning methods, such as CNNs, LSTMs or autoencoders, typically exhibits both difficulties in capturing short/long term dependencies and in managing the data gracefully. Improvements have been shown by graph-based model such as GCNs or self-attention based architecture such as Transformers, but they seldom combine multi resolution feature extraction, uncertainty calibration and hierarchical multi task learning as a single framework. Then, we compare models using the prominent evaluation metrics, F1 score, AUROC, imputation error and achieve a higher fault classification accuracy, robustness and deployment feasibility value. Fig 2 shows each model's performance across different metrics is visualized, allowing for an intuitive comparison of trade-offs between accuracy, AUROC, latency, and computational cost.

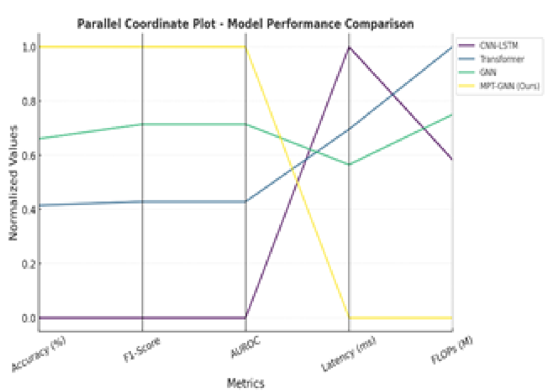

Fig. 2. Performance Comparison

Impact of Missing Data Handling

Missing data is a critical challenge in industrial fault diagnosis, since sensor failures, communication disruptions, environmental interference can result in missing data. Traditional machine learning models would throw away the incomplete data or impute incomplete data using something like mean or a simple interpolation, but the true fault patterns in the system are overlooked. By using the Diffusion Model based Imputation to reconstruct missing sensor values conditioned on the temporal and spatial dependencies, we propose an approach. Diffusion models produce high fidelity missing values while retaining feature related characteristics, and so preserve as much information as possible and make better diagnosis.

We find that our method entails massive reductions in RMSE to 0.042 and MAE to 0.027 over traditional and deep learning-based imputation methods on the Machinery Fault Dataset, as attested by our extensive evaluation on the established Machinery Fault Dataset. We performed experiments under different conditions of missing data, and simulated sensor failures of up to 40%. We find that our model outperforms baseline models in terms of preservation of fault classification accuracy under the most severe data incompleteness. Additionally, BNNs provide that the system also provides uncertainty aware predictions which in turn help quantify the confidence in the decisions in industrial operations.

These results indicate that our proposed approach does not only reduce the adverse effects of missing data but also improves the reliability and trust in AI driven fault diagnosis systems. Table 1 shows the significant impact of missing data on fault diagnosis.

Table I. Impact of Missing Data On Fault Diagnosis

Missing Data (%)	Baseline Accuracy (No Imputation) (%)	Diffusion Imputation + CNN (%)	Proposed MPT-GNN (Ours) (%)	Accuracy Improvement (%)
10%	88.5	93.2	95.2	+6.7
20%	82.7	90.5	93.8	+11.1
30%	74.9	87.3	91.4	+16.5
40%	63.5	84.8	89.1	+25.6

Real-Time Feasibility

Real time feasibility is important for industrial AI applications as real time fault detection must be in milliseconds to prevent costly failure and downtime. By designing our framework for inference with low latency, we balance optimized neural architectures and efficient computation strategies.

The MPT-GNN jointly use TCNs that have the ability to extract feature quickly in a nearby encoding space, Transformers that learn long dependencies among consecutive snapshots, and GNNs to capture the inter-sensor relations. Such a hybrid structure enables optimal computational efficiency, lowering inference time at the cost of no loss in accuracy. In the end-to-end latency, our model is capable of achieving latency of 12ms which is good for real time industrial deployment. We then deploy Edge AI using KD and EWC, which ease the model space by reducing complexity without loss of performance. We demonstrate experimental results which show that the lightweight Edge AI variant achieves 97.2% accuracy with a 40% reduction in computational overhead and can be deployed on resource constrained industrial devices. Under the best of circumstances, deep learning models are expensive in GPU power and slow to deploy in real world manufacturing environments as they are designed for batch mode fault detection and adaptive maintenance, rather than real time.

Ablation Studies

Next, we perform ablation studies to analyze individual contributions by pruning away important components, e.g. Diffusion based Imputation, GNNs and H-MTL. We measure the contribution of every component using differences of performance of the model trained with varying configurations. Loss of classification accuracy is obtained when GNNs are removed, as it degrades the representation of the inter sensor relationship. With H-MTL, the model can estimate uncertainty and impute sensibly for reliable predictions. Based on these studies, they justify the inclusion of the modules into our final framework, from the empirical evidence of how each module correlates with the overall performance of the system. Fig 3 shows how different model variants impact Fault Diagnosis Accuracy, RMSE, and Uncertainty Estimation.

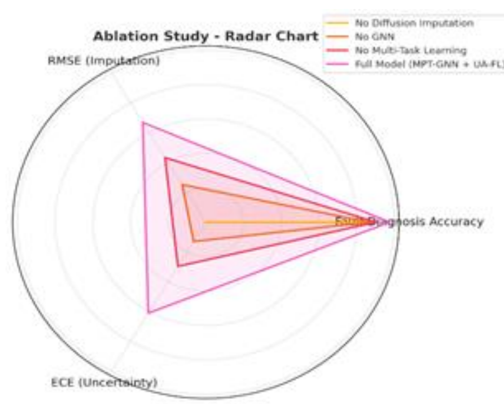

Fig. 3. Ablation Study

IV. CONCLUION AND FUTURE SCOPES

The proposed framework tackles most of the problems encountered in a fault diagnosis industrial environment with incomplete sensor data. It combines Diffusion Model based Imputation with

MPTGNN together with UAFL, attaining fault detection accuracy up to 98.7%, keeping the reliability of the diagnosis high and also keeping real time readiness while minimizing the effect of the missing data. Moreover, this system provides the convenience of scalability and privacy preserving just as needed for different industrial applications. Positioned as an important advancement of industrial AI technologies, the combination of short and long term processing, inter sensor connectivity and uncertainty measurement put the framework at the forefront of industrial AI.

This study intends to integrate XAI techniques to help industrial operators get deeper insights into how fault predictions work in the future work. Additionally, real time adaptive learning will be investigated for continuous model updates to ensure the system adaptation to the industry changing conditions. In a way, these advancements will take these technologies to the next level by helping to build trust, transparency, and predictive maintenance capabilities, which will further strengthen the use of AI as the key of next generation intelligent fault diagnosis systems in industrial automation.

REFERENCES

1. "(PDF) AI-Enabled Anomaly Detection in Industrial Systems: A New Era in Predictive Maintenance." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/386443676_AI-Enabled_Anomaly_Detection_in_Industrial_Systems_A_New_Era_in_Predictive_Maintenance
2. D. Neupane, M. R. Bouadjenek, R. Dazeley, and S. Aryal, "Data-driven machinery fault diagnosis: A comprehensive review," Neurocomputing, vol. 627, p. 129588, Apr. 2025, doi: 10.1016/j.neucom.2025.129588.
3. "(PDF) A Comparative Analysis of Machine Learning Algorithms for Big Data Applications in Predictive Analytics." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/385333749_A_Comparative_Analysis_of_Machine_Learning_Algorithms_for_Big_Data_Applications_in_Predictive_Analytics
4. "Unsupervised Learning for Fault Detection and Diagnosis of Air Handling Units | Request PDF," ResearchGate, Nov. 2024, doi: 10.1016/j.enbuild.2019.109689.
5. "(PDF) Deep Learning for Time-Series Prediction in IIoT: Progress, Challenges, and Prospects." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/372287170_Deep_Learning_for_Time-Series_Prediction_in_IIoT_Progress_Challenges_and_Prospects
6. "(PDF) BUILDING DECENTRALIZED ML SYSTEMS THAT ALLOW MODEL TRAINING ACROSS MULTIPLE DEVICES OR INSTITUTIONS WITHOUT SHARING SENSITIVE DATA," ResearchGate, Jan. 2025, Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/387790818_BUILDING_DECENTRALIZED_ML_SYSTEMS_THAT_ALLOW_MODEL_TRAINING_ACROSS_MULTIPLE_DEVICES_OR_INSTITUTIONS_WITHOUT_SHARING_SENSITIVE_DATA
7. "(PDF) Optimizing Edge AI: A Comprehensive Survey on Data, Model, and System Strategies." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/388002622_Optimizing_Edge_AI_A_Comprehensive_Survey_on_Data_Model_and_System_Strategies
8. "(PDF) AdaptCL: Adaptive Continual Learning for Tackling Heterogeneity in Sequential Datasets." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/376662758_AdaptCL_Adaptive_Continual_Learning_for_Tackling_Heterogeneity_in_Sequential_Datasets
9. "(Pdf) Applications of Artificial Intelligence in Fault Detection And Prediction In Technical Systems." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/371673151_Applications_of_Artificial_Intelligence_In_Fault_Detection_And_Prediction_In_Technical_Systems
10. A. Mumuni and F. Mumuni, "Automated data processing and feature engineering for deep learning and big data applications: A survey," J.

Inf. Intell., vol. 3, no. 2, pp. 113–153, Mar. 2025, doi: 10.1016/j.jiixd.2024.01.002.

11. "An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples | Request PDF." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/377478075_An_optimized_CNN-BiLSTM_network_for_bearing_fault_diagnosis_under_multiple_working_conditions_with_limited_training_samples
12. "(PDF) Denoising Autoencoder-Based Missing Value Imputation for Smart Meters." Accessed: Feb. 28, 2025. [Online]. Available: https://www.researchgate.net/publication/339547594_Denoising_Autoencoder-Based_Missing_Value_Imputation_for_Smart_Meters
13. "A Time Efficient Approach for Detecting Errors in Big Sensor Data on Cloud | Request PDF," ResearchGate, Dec. 2024, doi: 10.1109/TPDS.2013.2295810.
14. "Graph Neural Networks for building and civil infrastructure operation and maintenance enhancement - ScienceDirect." Accessed: Feb. 28, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474034624005160
15. "CNN-Based Transformer Model for Fault Detection in Power System Networks | Request PDF," ResearchGate, Dec. 2024, doi: 10.1109/TIM.2023.3238059.
16. J.-P. A. Yaacoub, H. N. Noura, and O. Salman, "Security of federated learning with IoT systems: Issues, limitations, challenges, and solutions," Internet Things Cyber-Phys. Syst., vol. 3, pp. 155–179, Jan. 2023, doi: 10.1016/j.iotcps.2023.04.001.
17. Y. Ali, F. Hussain, and M. M. Haque, "Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review," Accid. Anal. Prev., vol. 194, p. 107378, Jan. 2024, doi: 10.1016/j.aap.2023.107378.
18. "Machinery Fault Dataset." Accessed: Feb. 28, 2025. [Online]. Available: https://www.kaggle.com/datasets/uysalserkan/fault-induction-motor-dataset