



AI-Assisted Explainable Multimodal Predictive Maintenance of Industrial Bearings Using Vibration, Acoustic Analysis, Vision-Based Fault Detection, And Intelligent Maintenance Recommendation

¹Vansh Mittal

¹Mayo College, Ajmer

ABSTRACT

The emergence of industry 4.0 and smart manufacturing is resulting in the need for intelligent predictive maintenance systems that will help in the prevention of breakdowns and decrease maintenance costs. Even though vibration-based condition monitoring is commonly applied for diagnosis of bearing failure in industrial systems, it is faced by many challenges such as being incapable of diagnosing early stage and external faults. In this work, an AI-based multimodal predictive maintenance system that utilizes vibration analysis, acoustic analysis, thermal imaging and RGB vision-based inspection for precise, timely and non-invasive detection of faults is proposed. The performance of the hybrid CNN–LSTM model applied in the experiment was compared to SVM, Random Forest (RF), XGBoost, and traditional CNN. Contrary to conventional prediction maintenance methods which are focused on detection of faults, the proposed system utilizes an AI-assisted decision support system that will help in identifying the cause of the fault, the fault severity, and recommend appropriate maintenance strategies. The developed architecture had the accuracy of 98.2% and surpassed each of the individual sensors and traditional machine learning approaches. Vision-based AI algorithm identified the presence of thermal hot spots, lubricant leaks, bearing housing cracks, misalignment signals, and surface corrosion without taking apart the equipment while acoustic sensing increased the identification of early faults that had poor vibration signals. In general, the developed architecture is an intelligent approach to predictive maintenance using multiple sensors coupled with AI-based decision support system.

Keywords: Industry 4.0, Predictive Maintenance, Industrial Bearings, Artificial Intelligence, Vibration Analysis, Acoustic Analysis, Thermal Imaging, CNN–LSTM.

1. INTRODUCTION

The fast development of the industry 4.0 and smart manufacturing concepts has caused a radical change in the field of industrial maintenance due to the application of such advanced concepts as artificial intelligence (AI), IIoT, and sensing techniques. The implementation of the technique of predictive maintenance plays a key role in increasing the reliability of the machinery and decreasing both unexpected machine stoppages and maintenance expenses due to continuous machine state tracking. In spite of the widespread use of vibration analysis for the purpose of condition monitoring and bearing fault detection, there is the lack of detection of some early and external faults. Thus, in this research, the multimodal AI-assisted



predictive maintenance approach combining vibration, acoustic analysis, thermal, and RGB inspection will be implemented for bearing fault detection.

1.1 Industry 4.0, Bearing Fault Diagnosis, and Predictive Maintenance

The concept of Industry 4.0 has changed the face of manufacturing by incorporating artificial intelligence, Industrial Internet of Things, cloud computing, and cyber-physical systems within the operations of manufacturing industries. Bearings used in industry are important parts of rotating equipment and, when damaged, can lead to losses in production and higher maintenance cost. Traditional maintenance strategies, such as proactive and reactive maintenance, cannot always detect faults during the initial phase. Though vibration analysis is currently the most popular method of predicting problems in equipment, vibration-based systems can be impacted by background noise and inability to detect external faults. This calls for the development of advanced predictive maintenance systems.

1.2 Need for a Multimodal AI-Based Predictive Maintenance Framework

Some recent advances in artificial intelligence have made possible the use of multimodal sensing to diagnose faults intelligently. Acoustic sensing can pick up small changes in sounds that might signify the beginning of bearing deterioration, whereas thermal sensing picks up signs of overheating due to friction and lack of proper lubrication. In the same way, RGB camera-based sensing is capable of picking up visible faults, including oil leakages, cracks in the bearing housing, shaft misalignment, and corrosion without opening up the machine. The integration of vibration, acoustic, thermal, and visual information through a multimodal AI framework provides complementary diagnostic information, enhances fault detection accuracy, reduces false alarms, and supports real-time non-contact condition monitoring.

Despite the fact that vibration, acoustic, thermal, and computer vision have individually or together been explored for predictive maintenance, most of the existing approaches are mainly confined to fault detection and fault classification. Very few existing approaches provide interpretive AI-based decisions in terms of maintenance, fault severity estimation, or intelligent recommendations for maintenance purposes for assistance to maintenance engineers. Hence, it becomes essential to develop a predictive maintenance approach that will not only detect the fault without disassembling the machinery but will also interpret the fault detected, estimate its severity, determine the probable cause, and recommend maintenance accordingly.

1.3 Research Objectives

Objectives of this study are:

- To devise an AI-supported multimodal framework incorporating vibrations, sound, temperature, and RGB cameras for the diagnosis of faults in industrial bearings.
- To detect any bearing faults accurately and in real-time with the help of an AI-based hybrid CNN-LSTM model.
- To design an AI-supported decision-making system for fault interpretation and maintenance recommendations.

1.4 Novel Contributions

Contributions of this study include the following:



- An AI-based multimodal predictive maintenance approach that uses vibration, acoustic, thermal, and RGB camera data for predicting faults in bearings.
- Fault detection and classification in real time with the help of a hybrid model of CNN-LSTM without disassembling the machine.
- A decision support system based on AI that assists in fault interpretation, its severity evaluation, and intelligent maintenance planning.
- Smart planning of maintenance activities based on the fault condition predicted by the algorithm.
- Better accuracy and reliability than conventional vibration-based predictive maintenance systems used for Industry 4.0 applications.

2. LITERATURE REVIEW

Madhumitha et al. (2026) developed an image-based hybrid machine learning approach for bearing fault detection as a solution for predictive maintenance in industries and contributing to Sustainable Development Goal 9 (Industry, Innovation, and Infrastructure). Image-based data and hybrid machine learning approaches have been used in the research to identify the different bearing fault scenarios. The results showed that the application of image processing technology along with artificial intelligence can improve fault detection and allow condition monitoring. It was found that image-based intelligent systems can be highly beneficial for sustainable maintenance of industries.

Chen et al. (2026) conducted a thorough review on data-driven techniques and applications of artificial intelligence in reliability engineering and predictive maintenance. The review was focused on different machine learning, deep learning, and data analysis approaches for fault detection, diagnostics, and optimization of maintenance activities in industries. It is noteworthy that the review showed how intelligent predictive maintenance had increased the reliability of machines and fault prediction when compared with other approaches to maintenance. Nonetheless, there were several issues which made the application of intelligent maintenance difficult.

Chen et al. (2026) conducted an exhaustive review on artificial intelligence fault diagnosis technologies in ship rolling bearings and provided analysis of the latest trends, challenges, and future opportunities in the field. The authors reviewed both traditional methods of signal processing along with contemporary deep learning techniques applied for bearing health monitoring in various environmental conditions. According to the authors, vibration analysis has been the leading method but suffered from certain drawbacks associated with different environmental conditions and complexity of the faults. It has been pointed out that further studies should concentrate on combination of various sensor data and artificial intelligence technologies.

2.1 Research Gap

All the discussed studies revealed the applicability of artificial intelligence in detecting faults in bearings through image-based approach, vibration monitoring, acoustic analysis, thermal imaging, and data-driven predictive maintenance techniques. Nevertheless, most of the existing literature is concentrated more on the issue of fault detection and classification using



separate or combined types of sensors without giving much attention to the maintenance decision-making problem. While multimodal sensing has increased the precision of diagnostics, most of the systems lack the ability to provide the explanation for the fault cause, to assess its severity, to suggest the proper action in terms of maintenance, and to help maintenance engineers prioritize actions. Moreover, there has been no effort made towards creating a unified system capable of carrying out non-contact real-time detection of internal and external bearing faults without disassembling the machinery while giving intelligent assistance to maintenance engineers. Consequently, the proposed framework overcomes the problem and includes vibration signals, acoustic analysis, thermal imaging, and RGB camera inspection together with the mechanism of artificial intelligence-based decision-making to provide accurate diagnosis, fault interpretation, severity assessment, and maintenance recommendation.

3. MATERIALS AND METHODS

The research methodology utilized in this research is geared towards the development and analysis of an AI-based framework for predictive maintenance of industrial bearings based on their faults. This involves a systematic process that includes experimental set-up, collection of multimodal data, pre-processing of the collected data, development of AI models, and performance assessment.

3.1 Research Design

An experimental research methodology was used in this research to create and test an artificial intelligence-enabled multichannel predictive maintenance framework to diagnose faults in industrial bearings. In this framework, vibration detection, acoustic analysis, thermal imaging, and RGB camera-based inspection were used to identify faults in bearings without disassembling the machinery. The classification of various states of the bearing was done using a CNN–LSTM hybrid deep learning model.

3.2 Experimental Setup and Data Acquisition

The experimental data were collected on the basis of an industrial bearing test stand that was working under laboratory conditions. The test stand included an electric motor, shaft system, rolling bearings, and a load system that could be adjusted. Four different sensors were installed at the same time to collect experimental data: an accelerometer to measure vibrations, a high-sensitivity microphone to record acoustic signals, a thermal camera to monitor temperature distribution on the surface, and a visible camera to record visible machine states.

3.3 Bearing Health Conditions

Five bearing states were considered while conducting the experiments, namely healthy bearing state, inner race defect, outer race defect, rolling elements defect, and insufficient lubrication. Data were gathered for each of these operation states under the same speed and load conditions to keep consistency during training and evaluation of the models. The multimodal data set included vibration signals, acoustic signals, thermal images, and RGB images for each of these bearing states.

3.4 Data Preprocessing and Feature Extraction

The acquired vibrations and acoustic data were processed to eliminate the noise due to the environment as well as normalize the data prior to extracting features from them. The time domain and frequency domain features were extracted from the vibration data, while the acoustic data was transformed to its spectral form that can be used in deep learning. The thermal and RGB images were resized, normalized, and enhanced in order to improve their quality and enable feature extraction.

3.5 AI-Based Fault Classification

Classification of the combined multimodal features was done by implementing a combined CNN-LSTM model. The convolutional neural network (CNN) automatically extracted spatial information from thermal and RGB images, while the long short-term memory (LSTM) network learned temporal information from vibration and acoustic signals. The combined model classified the states of the bearings into healthy bearing, inner race fault, outer race fault, rolling element fault, and lack of lubrication. Comparison of the proposed model was carried out with SVM, RF, XGBoost, and conventional CNN models.

3.6 AI-Assisted Decision Support Module

The next step in the process after fault categorization was using an AI-enabled decision-making support system for improving maintenance planning. In accordance with the fault that is recognized by the CNN-LSTM system, the system determines the possible reason of the fault, evaluates the severity level of the fault, and generates maintenance suggestions.

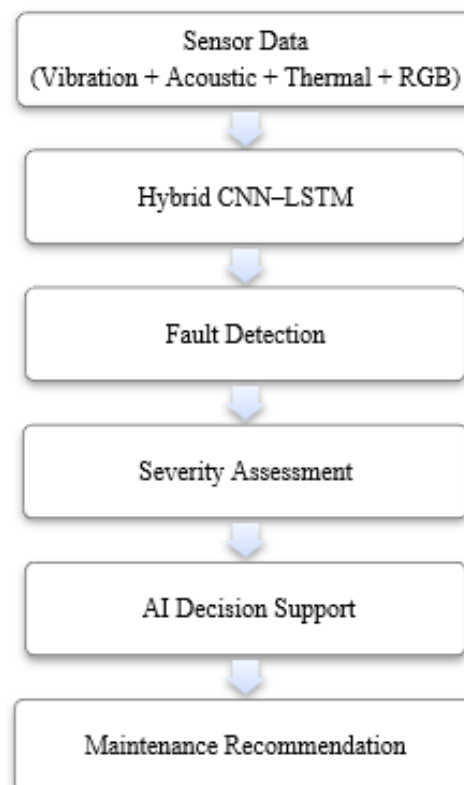


Figure 1: Workflow of the Proposed AI-Assisted Multimodal Predictive Maintenance Framework

3.7 Performance Evaluation

Performance metrics like Accuracy, Precision, Recall, F-Score, Processing Time, Detection Time, False Alarm Rate, and Detection Confidence were used to test the efficiency of the proposed predictive maintenance model. They were applied to test the diagnostic ability of the individual sensory methods as well as the suggested multi-modal AI model. The vision-based method was further tested for its detection ability to identify thermal hot spots, lubricant leaks, bearing housing fractures, misalignment indicators, and surface corrosion of machines without disassembling them.

3.8 Statistical Analysis

All experiments were carried out under the same operating conditions to guarantee the reproducibility of results. The model performances were analyzed through the use of classification metrics while comparisons were made to examine the performance of different machine learning algorithms. The results of the experiment were presented in tabulated and graphical form to enable an overall assessment of the proposed system.

4. RESULTS

This multivariate framework of predictive maintenance for AI-powered predictive maintenance was assessed based on the analysis of real-world data captured from industrial bearing test rigs when working in their normal as well as abnormal states. The data set used in this study included vibration, acoustic, thermal imaging, and RGB images obtained while the machinery was working continuously. Five bearing health conditions were considered in this study which included normal bearings, inner race defect, outer race defect, rolling elements defect, and lack of lubricant.

The findings from the experiment showed that the use of several sensing modalities increased the effectiveness of the fault diagnosis process as compared to using the vibration signals only. The inspection process based on the visual approach managed to identify problems including the thermal hot spots and leakage of the lubricants without having to disassemble the machinery. The use of the acoustic approach made it easier to detect faults whose vibration signals were very low.

Table 1: Performance of Individual Sensing Modalities for Bearing Fault Diagnosis

Sensing Modality	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Vibration Analysis	93.4	92.8	92.1	92.4
Acoustic Analysis	91.6	90.9	90.2	90.5
Thermal Imaging	89.8	89.1	88.6	88.8
RGB Camera Inspection	87.9	87.2	86.5	86.8
Proposed Multimodal AI Framework	98.2	98.0	97.8	97.9

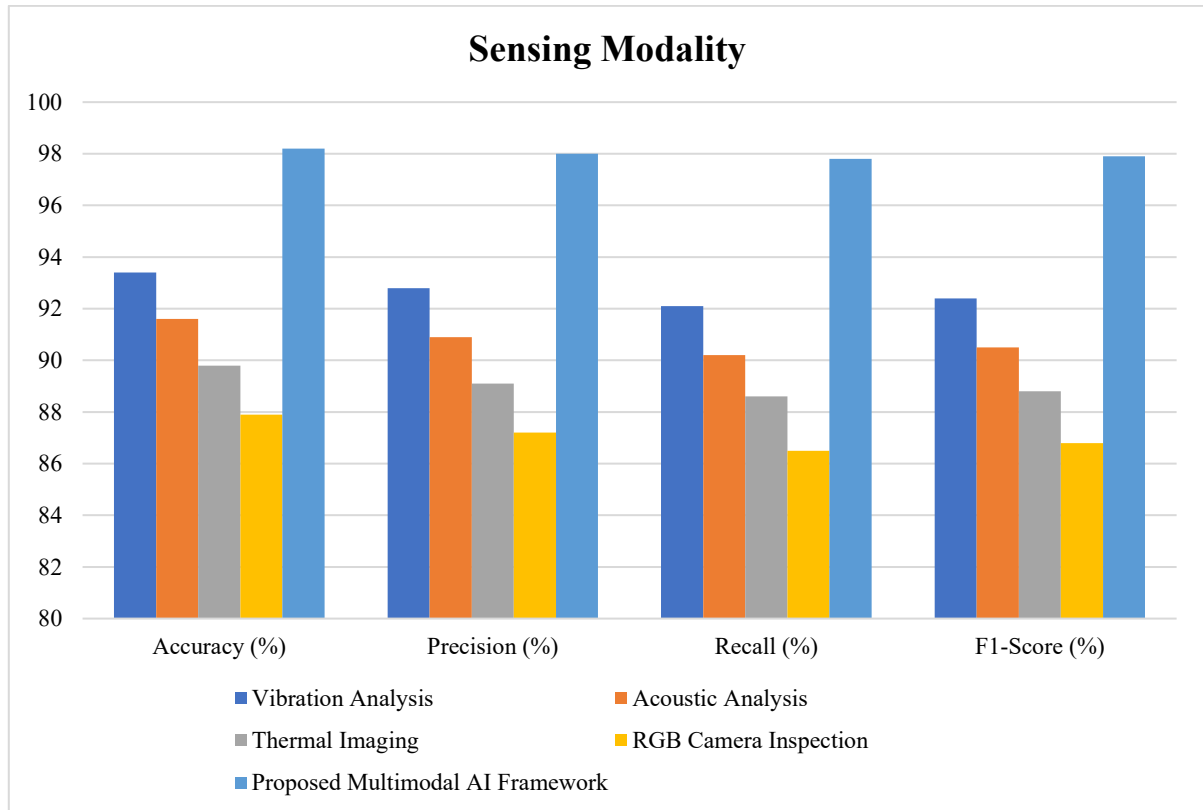


Figure 1: Graphical Representation of Performance of Individual Sensing Modalities for Bearing Fault Diagnosis

Table 1 shows the diagnostics capabilities of individual sensing technologies and the proposed multimodal AI model. Vibration sensing gave the best result among individual sensor technologies (93.4%), followed by acoustic sensing (91.6%), thermal imaging (89.8%), and RGB camera testing (87.9%). However, combining vibration, acoustic, thermal, and visual data greatly increased diagnostic capacity, allowing the proposed multimodal model to obtain a classification accuracy rate of 98.2%. The results reveal the high effectiveness of using sensor fusion to eliminate weaknesses of individual sensors.

Table 2: Detection Performance for Different Bearing Fault Types Using the Proposed Multimodal AI Model

Bearing Condition	Detection Accuracy (%)	Average Detection Time (s)	False Alarm Rate (%)
Healthy Bearing	99.3	0.42	0.4
Inner Race Defect	98.7	0.48	0.7
Outer Race Defect	98.1	0.50	0.8
Rolling Element Defect	97.9	0.53	0.9
Lubrication	96.8	0.46	1.2

Deficiency			
------------	--	--	--

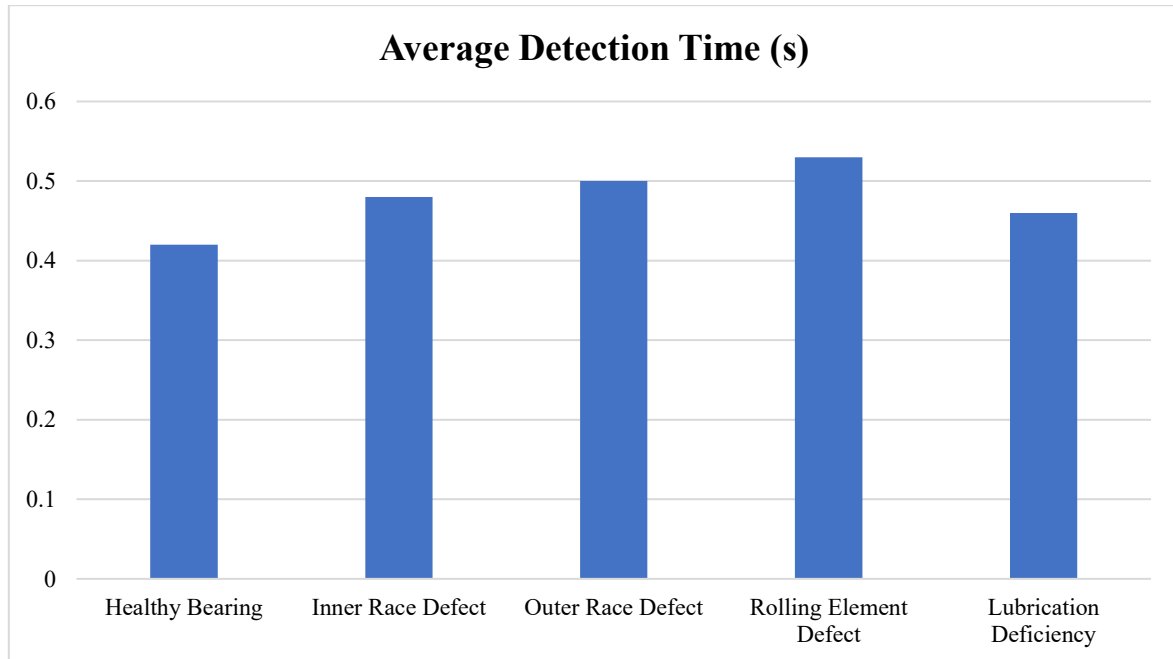


Figure 2: Graphical Representation of Detection Performance for Different Bearing Fault Types Using the Proposed Multimodal AI Model

The performance results of the suggested multi-modal approach according to fault type are provided in Table 2. The proposed model precisely recognized the healthy condition of bearings with 99.3% accuracy and a very low rate of false alarm (0.4%). In faulty situations, the highest accuracy of detection was reached in the case of inner race defects (98.7%), but in case of lubrication shortage, lower accuracy (96.8%) was obtained since its features emerged progressively. However, all kinds of faults were detected within 0.55 s.

Table 3: Comparative Performance of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Processing Time (ms)
Support Vector Machine	91.8	91.0	90.4	90.7	96
Random Forest	93.7	93.2	92.5	92.8	88
XGBoost	95.1	94.8	94.2	94.5	82
CNN	96.4	96.0	95.8	95.9	71
CNN-LSTM	98.2	98.0	97.8	97.9	63

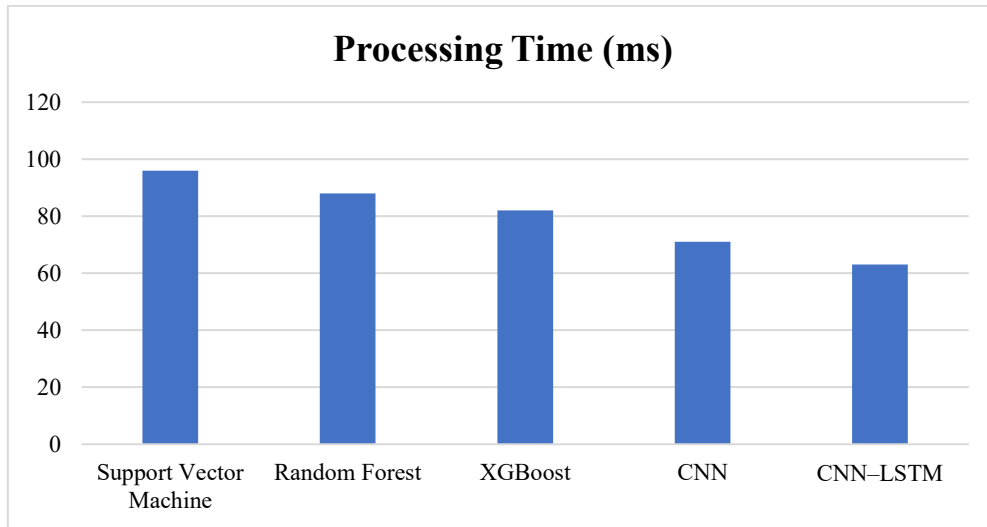


Figure 3: Graphical Representation of Comparative Performance of Machine Learning Models

The results for the comparative analysis of the proposed CNN-LSTM technique against conventional machine learning algorithms are provided in Table 3. Although Support Vector Machine and Random Forest were capable of providing adequate results in classification, they were inferior to deep learning methods in terms of diagnostic accuracy. The XGBoost algorithm enhanced the results in terms of classification, yet the proposed architecture was superior to all other models in terms of accuracy (98.2%) and minimum processing time (63 ms).

Table 4: Vision-Based Non-Invasive Inspection Results

Detected Condition	Detection Accuracy (%)	Average Confidence Score (%)	AI Detection Without Machine Disassembly
Thermal Hotspots	98.5	98.2	Yes
Lubricant Leakage	97.8	97.3	Yes
Bearing Housing Crack	96.9	96.4	Yes
Shaft Misalignment Indicators	97.2	96.8	Yes
Surface Corrosion	95.8	95.2	Yes

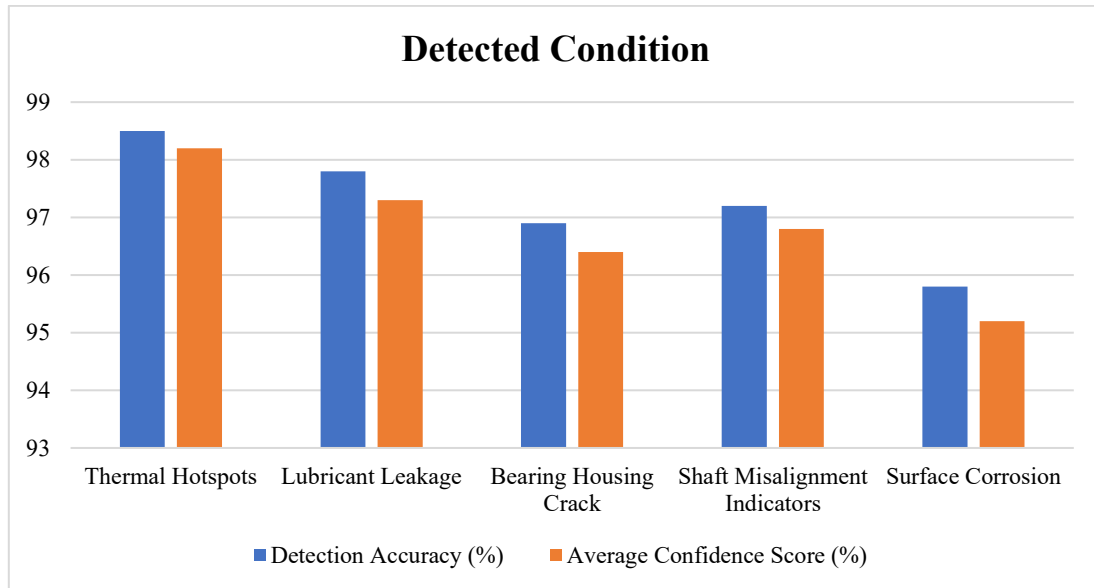


Figure 4: Graphical Representation of Vision-Based Non-Invasive Inspection Results

The ability of the AI inspection system based on vision technology to recognize mechanical defects through visual identification without the need to take apart the industrial machinery is shown in Table 4. The thermal anomalies were recognized most effectively (98.5%), while lubrication leaks and shaft misalignments were identified with accuracies of 97.8% and 97.2%, respectively. Bearing housings with cracks and corrosion could also be accurately detected (accuracy > 95%). It can therefore be seen that the integration of thermal and RGB cameras with AI results in effective inspection systems.

Table 5: AI-Assisted Decision Support and Maintenance Recommendations

Detected Fault	Severity Level	AI-Assisted Maintenance Recommendation
Lubrication Deficiency	Moderate	Re-lubricate the bearing and inspect the lubrication system within 24 hours.
Inner Race Defect	High	Schedule bearing replacement during the next planned maintenance shutdown.
Outer Race Defect	High	Inspect bearing alignment and replace the damaged bearing if necessary.
Bearing Housing Crack	Critical	Stop machine operation immediately and replace the damaged component.
Thermal Hotspot	Medium	Inspect the lubrication and cooling systems to prevent overheating.

The results presented in Table 5 illustrate the feasibility of the proposed AI-enabled decision support system besides fault diagnosis. According to the type of detected fault, the severity

level is estimated and maintenance measures are recommended accordingly. Insufficient lubrication was found to be a medium fault which requires quick re-lubrication. The defect in the inner and outer races was found to be a high-severity fault which needs replacement or realignment. Cracks in the bearing housing were considered to be a critical fault, which required immediate shutdown of the machine. The thermal hotspot was given a medium severity rating, requiring an investigation into the lubrication/cooling systems. This highlights the ability of the proposed framework to help maintenance engineers prioritize maintenance operations and reduce downtime.

5. DISCUSSION

The findings indicated that the suggested multimodal AI-assisted framework for predictive maintenance offered an efficient way to improve the performance of industrial bearing fault diagnosis through the use of vibration sensing, acoustic signal processing, thermal sensing, and visual inspection. In comparison to the other sensing approaches, the highest classification accuracy (98.2%) was obtained through the suggested framework, which had higher precision, recall, and F1 score as well. The lowest false alarm rate and fast detection time made this clear.

The hybrid CNN-LSTM algorithm performed better than traditional machine learning algorithms, while the vision-based AI algorithm was able to identify the presence of thermal hotspots, lubrication leakage, bearing housing damage, misalignment of the shaft, and surface corrosion of the machinery without disassembling the system. In contrast to traditional predictive maintenance systems that mainly focused on fault detection, the proposed approach also includes a decision-making module powered by artificial intelligence that helps in interpreting the condition of the fault, assessing its severity, and determining necessary actions for maintenance. The proposed approach provides a smart way for predictive maintenance in the environment of Industry 4.0.

6. CONCLUSION

The findings indicate that the proposed artificial intelligence (AI) based multi-modal predictive maintenance framework effectively improved the diagnosis of industrial bearing fault with the use of vibration analysis, acoustic signal processing, thermal imaging, and vision-based inspection. The hybrid CNN-LSTM framework attained the best classification accuracy (98.2%) with minimal false alarms and fast detection rate, compared to other machine learning techniques. Additionally, the AI-based vision framework was effective in detecting thermal hot spots, lubrication leakage, bearing housing cracking, shaft misalignment, and corrosion without disassembling the machinery. As opposed to other predictive maintenance frameworks that mainly focus on detecting faults, the proposed AI-based predictive maintenance framework has also been developed in such a way that it not only detects faults but also interprets them in order to estimate their severity and suggest appropriate actions.

REFERENCES

1. Madhumitha, S., Narayana, S. K. S., Dhevi, S. Y., Venkataraman, M., & Nadarajah, S. (2026). A Hybrid Machine Learning Framework for Image-Based Bearing Fault



- Diagnosis Advancing SDG Goal 9: Sustainable Industry and Innovation. Array, 101017.
2. Chen, X., Li, W., Tianze, X., Ruizhi, O., & Kaiye, G. (2026). Data-driven methods and artificial intelligence in reliability and maintenance: A review. *Mathematics*, 14(5), 899.
 3. Chen, R., Luo, Q., Qin, Y., Yang, L., Huang, Y., & Zhou, H. (2026). A comprehensive review of fault diagnosis methods for ship rolling bearings under the AI revolution: challenges and opportunities. *Measurement Science and Technology*, 37(23), 232002.
 4. Rzeszowska, A., Błażej, R., & Jurdziak, L. (2025). Non-Destructive Testing for Conveyor Belt Monitoring and Diagnostics: A Review. *Applied Sciences*, 15(24), 13272.
 5. Wang, Y., Yu, W., Yu, C., Shi, H., & Li, W. (2026). A survey on multimodal learning for industrial diagnostics: a data dimensionality perspective. *Artif. Intell. Rev.*, 59(5), 126.
 6. Mahmood, M. S., & Shareef, I. R. (2024). Applications of Artificial Intelligence for Smart Conveyor Belt Monitoring Systems: A Comprehensive Review. *Journal Européen des Systèmes Automatisés*, 57(4), 1195.
 7. Mattaparthi, R. (2023). Connected Fleet Intelligence: Edge-Centric Analytics and Computer Vision for Predictive Manufacturing and Asset Resilience. *International Journal of Advanced Research in Computer Science & Technology (IJARCST)*, 6(5), 9077-9088.
 8. Ahangar, M. N., Farhat, Z. A., Sivanathan, A., Ketheesram, N., & Kaur, S. (2026). Explainable AI-driven quality and condition monitoring in smart manufacturing. *Sensors*, 26(3), 911.
 9. Hassan, M., Kirwin, R., Rakurty, C. S., & Mahajan, A. (2026). Advanced Sensor Technologies in Cutting Applications: A Review. *Sensors*, 26(3), 762.
 10. Dos Santos, L. F. D., Canuto, J. L. D. S., Aylon, L. B. R., Mariani, V. C., & De Souza, R. C. T. (2026). A Multimodal Model for Fault Diagnosis in Induction Motors Using Thermal Images and Large Language Models. *IEEE Access*.
 11. Panda, A. K., Rao, D. C., & Sahoo, G. K. (2025). Multi-Domain Feature Learning with Transformer-Derived LLMs for Intelligent Fault Diagnosis. *MESSAGE*, 156, 62.
 12. Fu, H., Rao, J., Deng, F., Wang, Y., Zhao, B., Liu, Z., ... & Xu, L. (2025). AIoT: Artificial Intelligence and the Internet of Things for monitoring and prognosis of systems and structures. *IEEE Transactions on Instrumentation and Measurement*.
 13. Bris-Peñalver, F. J., Verdecia-Peña, R., & Alonso, J. I. (2026). A Survey of AI-Enabled Predictive Maintenance for Railway Infrastructure: Models, Data Sources, and Research Challenges. *Sensors*, 26(3), 906.
 14. Golovastikov, N. V., Kazanskiy, N. L., & Khonina, S. N. (2025, June). Optical fiber-based structural health monitoring: advancements, applications, and integration with



- artificial intelligence for civil and urban infrastructure. In Photonics (Vol. 12, No. 6, p. 615). MDPI.
15. Nguyen, T. Q., Phan-Vu, P., & Nguyen, P. T. (2026). AI-based damage detection in prestressed concrete beams: a vision-integrated deep learning framework for crack localization and severity classification. *Advances in Bridge Engineering*, 7(1), 6.
 16. Pouloupoulos, G., & Avramopoulos, H. (2026). Review of Optical Fiber and Integrated Photonic Sensors for Industry and Smart Manufacturing: Technologies, Applications, Structural Health Monitoring and AI-Enabled Sensing. *Sensors (Basel, Switzerland)*, 26(11), 3581-3581.
 17. Soleimanpour, R. (2026). Recent Advances in AI and Signal Processing for PZT-Based Structural Health Monitoring. *Infrastructures*, 11(7), 228.
 18. PAL, H. (2025). AI-Augmented FEA for Offshore Wind Turbine Equipment: A State-of-the-Art Methodology for Accelerated Fabrication Review.
 19. Abbas, S. Q. (2026). AI-Driven Structural Health Monitoring and Predictive Maintenance of Bridge Infrastructure Using IoT Technologies.
 20. JAYARAM, M. (2026). ARTIFICIAL INTELLIGENCE AND APPLICATIONS IN CORE ENGINEERING. PHI Learning Pvt. Ltd.