Proceedings of the 11th North American International Conference on Industrial Engineering and Operations Management, Milwaukee, Wisconsin, USA, June 9-11, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260212
Published: June 9, 2026

Multichannel Acoustic Feature Fusion for Predictive Maintenance in High-Speed Robotic Manufacturing

Md Omar Al Javed
Mechanical and Nuclear Engineering
Tennessee Technological University
Cookeville, TN, USA
mjaved42@tntech.edu

Ayantha Senanayaka
Manufacturing and Engineering Technology
Tennessee Technological University
Cookeville, TN, USA
asenanayaka@tntech.edu

Abstract

High-speed robot-based manufacturing (RBM) systems require reliable and efficient maintenance strategies to minimize downtime and ensure operational continuity. Traditional predictive maintenance approaches rely on intrusive and costly sensing devices, such as vibration and torque sensors, which increase installation complexity and operational overhead. As a non-invasive alternative, acoustic sensing enables continuous condition monitoring. Current acoustic monitoring studies primarily depend on traditional single-sensor time-domain analysis and multisensory approaches that involve data-level fusion strategies. These methods often fail to adequately capture the complex spatial dynamics and noisy industrial sounds present in robotic environments. To overcome this challenge, this study introduces a Multichannel Acoustic Feature Fusion (MCAFF) framework to improve robotic condition identification for predictive maintenance. The proposed approach integrates signals from spatially distributed microphones to generate fused time-frequency representations using the Short-Time Fourier Transform (STFT). These representations are then used to train a Convolutional Neural Network (CNN) for classifying robot operating conditions. Experimental validation was conducted on a FANUC M-1iA/0.5A robot performing high-speed tasks at varying speeds under constant payload conditions. Results demonstrate that the proposed MCAFF framework improves classification accuracy by 15% compared to single-sensor approaches and by 10% compared to conventional data-level fusion methods. The findings highlight the effectiveness of multichannel acoustic feature fusion as a scalable, non-invasive solution for improving predictive maintenance and reliability in RBM systems.

Keywords
Robot-based manufacturing, predictive maintenance, multichannel acoustic sensing, feature fusion, convolutional neural networks.

1. Introduction
High-speed robot-based manufacturing (RBM) systems are increasingly utilized in precision-critical industrial environments. Ensuring reliability and minimizing downtime in these systems is essential for maintaining productivity and cost efficiency (Soori et al. 2024). To meet this need, predictive maintenance has emerged as a key strategy for detecting faults before they lead to failures (Kumar et al. 2018). Transitioning from reactive repairs to proactive interventions effectively extends equipment lifespan and reduces operational disruptions. Traditional predictive

maintenance methods rely on intrusive measurement devices, such as vibration sensors, torque meters, and temperature gauges, which are installed on robots to collect real-time operational data (Hassan et al. 2024). By analyzing this data, maintenance teams can identify patterns or anomalies indicating potential failures. While effective, these methods often involve significant installation downtime, expensive sensing devices, and skilled personnel for accurate interpretation (Achouch et al. 2022). Attaching sensors to moving robotic joints can disrupt operational dynamics and increase the risk of resource mismanagement. This risk is particularly significant in complex RBM systems with frequent interacting components, where improper sensor integration may negatively impact overall efficiency (Qiao and Weiss 2017).

As a non-invasive alternative, acoustic sensing addresses these limitations by enabling continuous condition monitoring without modifying the physical structure of robotic systems (Gong et al. 2025). Existing acoustic-based diagnostic methods primarily rely on single-channel time-domain analysis and conventional machine learning (ML) models (Ntalampiras 2021). This reliance limits their ability to capture complex operational dynamics and makes them exposed to non-stationary noise and overlapping sound sources, which are common in industrial robotic environments (Qurthobi et al. 2022).

Recent research has demonstrated that integrating multiple sensing sources significantly enhances detection capabilities compared to traditional single-source sensing methods (Tang et al. 2023). This approach provides a more comprehensive understanding of complex environments and systems. However, current multisensor technologies mainly rely on time-domain data-level fusion techniques that focus on synchronizing and merging data collected from different sensors over time. This method can introduce additional noise into the datasets and complicate model development tasks. Furthermore, there remains a considerable gap in the application of these technologies, especially concerning high-speed robot maintenance. Most existing research and development efforts have focused on general machine maintenance issues, overlooking the unique challenges and complexities of maintaining high-speed RBM systems (Wang et al. 2021).

Identifying the operating conditions of robots at high speeds in fast-paced environments is crucial. There is a significant need for advanced methodologies that ensure physical consistency across multiple channels, effectively capture nonlinear interactions across various acoustic channels, and extract meaningful features from complex robotic settings. To address these challenges, this study proposes a Multichannel Acoustic Feature Fusion (MCAFF) framework for identifying operational speeds in RBM by integrating spatially distributed acoustic information. The framework enhances the ability to discriminate between different robot operating conditions.

The proposed method conducts feature-level fusion by first transforming raw acoustic signals from spatially distributed microphones into time-frequency representations using the Short-Time Fourier Transform (STFT). These spectrograms are then combined to create a unified multi-channel tensor that preserves both the spatial and spectral characteristics of the acoustic emissions. This resulting representation serves as input to Convolutional Neural Networks (CNNs), allowing the extraction of highly discriminative features and enabling classification. The proposed method's effectiveness was validated through experiments aimed at identifying conditions in high-speed robots. These experiments employed the FANUC M-1iA/0.5A high-speed industrial robot manipulator, specifically designed for rapid and precise tracing under multi-speed conditions.

This research makes the following technical contributions:

1. Develop a multichannel acoustic sensing framework for robotic condition identification, improving operating condition detection performance using feature fusion techniques.
2. Evaluate the effectiveness of the proposed MCAFF framework against single-sensor baselines and conventional data-level fusion methods.
3. Demonstrate the applicability of the proposed MCAFF approach in condition identification in high-speed RBM environments.

The remainder of the paper is organized as follows: Section 2 reviews current methodologies for RBM, vibration and acoustic-based monitoring, multisensor fusion, and identifies challenges and research gaps. Section 3 presents the proposed methodology. Section 4 details the experimental setup and data collection. Section 5 presents results, validation, discussion, and future work. Finally, Section 6 summarizes the conclusions of this study.

2. Literature Review
The RBM systems are now integral to modern industrial automation, offering unparalleled precision and efficiency. However, the continuous operation of these systems at high speeds exposes mechanical components to significant stress, making them susceptible to sudden failures (Soori et al. 2024). Consequently, maintaining the reliability of these systems is critical, as unexpected downtime can lead to severe production bottlenecks and financial losses.

2.1 Predictive Maintenance in RBM
Predictive maintenance has become a cornerstone of modern RBM, particularly within the framework of Industry 4.0. The primary goal of this approach is to ensure system reliability and minimize unexpected downtime by assessing the health status of robots before failures occur. In early foundational work, Mobley (2002) established the principles of condition-based maintenance and initiated the shift from reactive to proactive strategies. Peng et al. (2010) categorized prognostic models to underscore the growing importance of data-driven approaches. Achouch et al. (2022) provided a comprehensive overview of predictive maintenance in Industry 4.0, noting that the integration of Internet of Things (IoT) sensors and advanced analytics has fundamentally transformed industrial condition monitoring. In the specific context of high-speed robotics, predictive maintenance strategies must account for rapid operational cycles and complex joint kinematics, which complicate fault detection compared to traditional stationary machinery (Uhlmann et al. 2020). In another study, Bogue et al. (2025) show that the implementation of condition monitoring on a continuous basis is essential for detecting early faults in industrial robots, thereby preventing catastrophic failures during high-speed operations.

2.1.1 Vibration-based Monitoring in RBM
Historically, vibration-based monitoring has been the most prevalent approach for RBM. Jardine et al. (2006) extensively detailed the effectiveness of vibration analysis in detecting mechanical faults. This finding has since been reinforced by recent comprehensive reviews (Lei et al. 2020; Tama et al. 2023). While these methods provide accurate fault detection, they require physical sensor integration. This direct contact increases system complexity and deployment costs, particularly in RBM systems. Moreover, the physical attachment of vibration sensors to moving robotic joints can interfere with high-speed operational dynamics and increase the risk of resource damage (Makulavičius et al. 2023). To overcome these limitations, alternative non-invasive techniques have been explored.

2.1.2. Acoustic-based Monitoring in RBM
Acoustic emission techniques have gained significant attention due to their non-invasive nature. For instance, Qurthobi et al. (2022) highlighted the effectiveness of acoustic analysis in detecting anomalies in mechanical systems without requiring dedicated contact sensors. In the context of RBM, acoustic monitoring presents unique opportunities and challenges. Gong et al. (2025) systematically reviewed acoustic-based fault detection for robotic arms, emphasizing its importance in industrial automation where non-invasive and real-time diagnostics are essential. Prolonged high-speed operation of robotic arms frequently leads to mechanical degradation, making acoustic monitoring a vital tool for early fault detection without disrupting the manufacturing process (Gong et al. 2025). To address the challenge of overlapping background noise, Senanayaka et al. (2024) demonstrated that isolating fault sources from complex mixtures of industrial sounds prior to ML-based classification significantly enhances diagnostic accuracy. Despite these advancements, most studies rely on single-sensor configurations. This reliance limits spatial awareness and makes the systems susceptible to the non-stationary noise typical of factory floors.

2.2. Sensor Fusion Methods
To improve prediction accuracy and robustness against environmental noise, multisensor data fusion has been proposed. As Tsanousa et al. (2022) emphasized data fusion techniques effectively address multimodality by combining heterogeneous sensor data, with early-stage feature fusion being particularly prevalent in industrial prognosis. In the context of advanced manufacturing, Lee et al. (2025) demonstrated that fusing multi-directional scanning data significantly enhances defect detection accuracy, highlighting the broader applicability of spatial data integration. Furthermore, recent studies emphasize that the level at which fusion occurs significantly impacts performance (Li et al. 2024; Xiao et al. 2025). Advanced feature fusion methods have been shown to overcome the structural limitations of raw data-level fusion, leading to significantly higher classification accuracy in complex machinery (Wang et al. 2021). Also, researchers showed that multi-channel sensor fusion using frequency-domain multilinear principal component analysis effectively captures complex fault signatures in rotating machinery (Al Mamun et al. 2023; Senanayaka et al. 2024).

2.3 Challenges and Research Gaps

Existing literature emphasizes the need for non-invasive, scalable, and highly accurate predictive maintenance solutions for RBM systems. However, current fusion techniques primarily rely on simple concatenation and analyses in either the time or frequency domain, along with late-stage decision-level or data-level fusion. Most studies have focused on rotating machinery rather than RBM systems. Furthermore, these approaches may not fully leverage the complementary spatial information that distributed sensors provide during the feature extraction stage. The integration of multichannel acoustic data with early-stage feature fusion remains underexplored in RBM. To address these limitations, this study introduces the MCAFF framework.

3. Methods

The proposed MCAFF framework combines multichannel acoustic signals with deep learning techniques for condition classification. Acoustic signals are first collected from several microphone arrays strategically positioned around the RBM system. Figure 1 illustrates the workflow of the proposed MCAFF methodology.

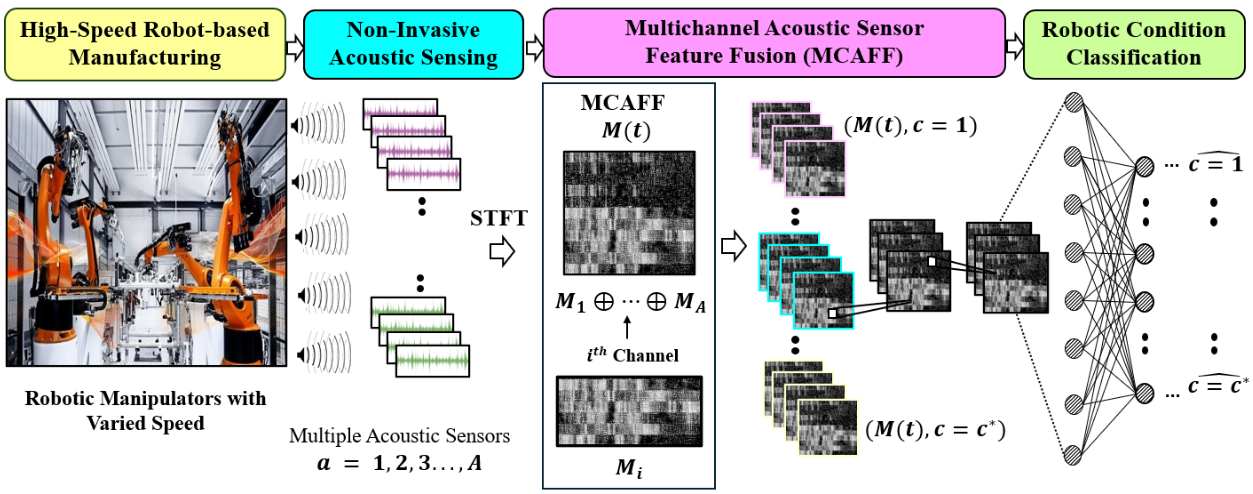

Figure 1. Workflow diagram of the proposed MCAFF method

3.1 Developing MCAFF Representation using Spatially Located Acoustic Sensors

Let the raw acoustic data consist of multiple time-domain signals collected from a set of sensor channels, denoted as $a = 1,2, \ldots, A$, where A represents the total number of channels. Each channel corresponds to a microphone that captures spatial acoustic information from the RBM system and its surrounding environment. In this study, each microphone operates as a single (mono) channel. At a given time, instant t, the discrete time-domain signals collected from all channels can be expressed as:

$$X(t) = \{x_1(t), x_2(t), \ldots, x_A(t)\} \tag{1}$$

where $x_i(t)$ represents the signal acquired from the i-th channel. The operational condition of the RBM system is characterized using acoustic signatures within a spatial domain S, where each $s \in S$ denotes a specific system condition. These conditions may include robotic operation at given speed, mechanical wear, bearing faults, motor malfunctions, lubrication deficiencies, electrical anomalies, and environmental noise. Suppose that each time-domain signal is associated with a condition label $c \in C$, where C represents the set of all possible condition categories identified in RBM. A labeled dataset can therefore be represented as:

$$\mathcal{D} = \{(X(t), c)\} \tag{2}$$

where $X(t)$ is the multi-channel signal and c is the corresponding condition label. To extract informative features, each time-domain signal is transformed into a time-frequency representation using the Short-Time Fourier Transform (STFT). This operation is denoted as:

$$F_i = f(x_i(t)) \quad (3)$$

where $f(\cdot)$ represents the STFT operation applied to the i-th channel. The resulting spectral features are standardized to a uniform size U for all channels. The multi-channel feature representation is then constructed by combining the features from all channels:

$$M = M_1 \oplus M_2 \oplus \cdots \oplus M_A \quad (4)$$

where M_i denotes the feature matrix of the i-th channel and $\oplus$ represents the fusion operation. The fused feature tensor at time t is defined as:

$$M(t) \in \mathbb{R}^{U \times D} \quad (5)$$

where D represents the total feature dimension after fusion across all channels. Finally, the tensor $M(t)$ can be interpreted as a two-dimensional matrix:

$$M(t) \in \mathbb{R}^{l \times z} \quad (6)$$

where l and z correspond to the row and column dimensions, respectively. Each element in the matrix represents a numerical feature value, analogous to a pixel intensity in an image. These representations, generated across time and operating conditions, provide discriminative features for classification and predictive modeling.

3.2 Condition Classification Using CNN

The feature representation $M(t)$, obtained from the proposed MCAFF framework, is utilized as the input to a CNN for condition classification. CNNs are widely used for image-based classification due to their ability to learn complex spatial relationships within structured data. In this study, $M(t) \in \mathbb{R}^{l \times z}$ is treated as an image-like representation. Through convolutional operations, CNN extracts hierarchical features that capture discriminative patterns associated with different operating conditions. This capability is particularly important for RBM condition classification, where multiple condition types may occur, making accurate diagnosis more challenging. By leveraging spatial correlations in the MCAFF representation, CNN can effectively distinguish between both complex conditions.

A typical CNN architecture consists of convolutional layers, pooling layers, nonlinear activation functions, fully connected layers, and an output layer. The network is designed to classify input samples into condition categories $c \in C$, consistent with the labeling defined in Equation (2). The CNN maps $M(t)$ to a feature vector $x \in \mathbb{R}^D$ through its learned layers. The probability of assigning the input to class $c_k \in C$ is computed using the Softmax function:

$$P(c_k \mid M(t)) = \frac{exp\left(w_{c_k}^T x + b_{c_k}\right)}{\sum_{i=1}^{K} exp\left(w_{c_i}^T x + b_{c_i}\right)} \quad (7)$$

where $P\left(c_k \mid M(t)\right)$ denotes the predicted probability of class c_k, w_{c_k} and b_{c_k} are the weight vector and bias associated with class c_k, respectively, and $K = \mid C \mid$ represents the total number of condition classes. The predicted class label is determined by selecting the class with the highest posterior probability. The CNN model is trained by minimizing the categorical cross-entropy loss between the predicted probabilities and the true class labels.

4. Experimental Setup and Data Collection

This section presents the details of data description and experimental results for the proposed method validation. The experiments focus on a RBM system's multiple condition classification.

4.1. Robotic System and Operating Conditions

The experiments used the FANUC M-1iA/0.5A high-speed industrial RBM, known for its advanced parallel (delta-type) manipulation. Designed for rapid pick-and-place tasks and precision tracing, it excels in industries requiring quick, accurate movements. With 6 degrees of freedom (DOF), the robot offers exceptional agility and precision. It is controlled by an R-30iB FANUC controller for easy integration and programming. In this study, the robot was programmed to perform a linear tracing loop under controlled conditions. A constant payload of 40 g was attached to the end-effector during all trials to simulate realistic operational conditions (Figure 2).

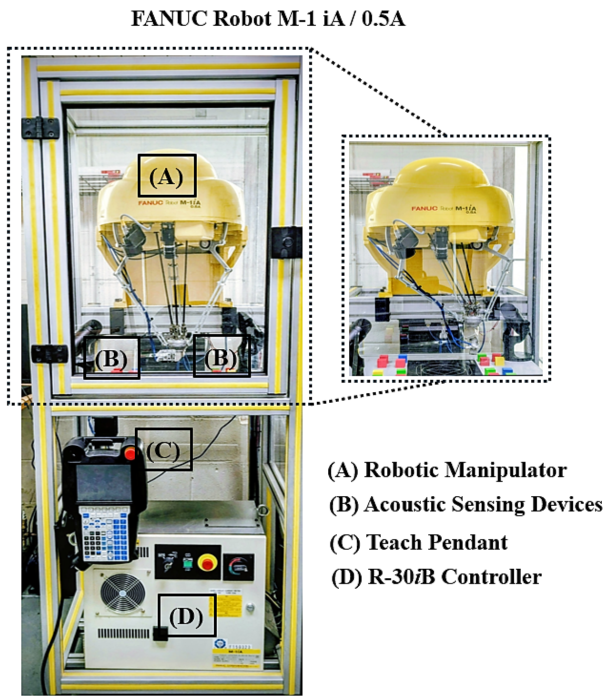

Figure 2. Experimental setup for identifying conditions using high-speed robotic operations

Three distinct operational parameters have been established based on the robot's linear velocity, each promoting specific dynamic characteristics that are essential for the classification task.

1. Condition 1: 200 mm/s - Lower velocity setting enables high-precision movements, allowing the robotic system to perform complex trajectories while maintaining exceptional positional accuracy.

2. Condition 2: 500 mm/s - Operating at this intermediate velocity strikes an optimal balance between speed and precision, facilitating efficient task execution while ensuring accuracy in motion tracking.

3. Condition 3: 1000 mm/s - High-speed configuration demonstrates the robot's capability for the rapid execution of tracing tasks, effectively showcasing its performance in time-sensitive applications where velocity is of utmost importance.

These varying conditions not only represent the robotic system's versatility but also serve as distinct classes in the classification task, enabling an in-depth analysis of the robot's performance under different dynamic influences. These conditions represent varying dynamic behaviors of the RBM system and are treated as distinct classes $c \in C$in the classification task.

4.2. Acoustic Data Acquisition and Data Set Construction

In our study, we used two high-quality condenser microphones positioned strategically around the base of the robot to capture acoustic signals from various locations. The setup featured a dual-channel configuration with a sampling frequency of 44.1 kHz, ensuring detailed recordings. Each trial lasted 3 minutes, and we conducted 5 independent trials per condition for reliability. All signals were recorded simultaneously, maintaining temporal synchronization, which was essential for analyzing sound variations in the environment.

The collected acoustic signals were segmented to form labeled samples. For each condition $c \in C$, approximately 180 samples per trial were generated, resulting in a comprehensive dataset. The dataset is defined as Equation 2,

$X(t)$ represents the multi-channel acoustic signal and c denotes the corresponding operating condition label. Each time-domain signal $x_i(t)$ was transformed into a time-frequency representation using the STFT, as defined in Equation (3). The STFT parameters are window size: 1024, with hop length 512. The resulting spectral features from each channel were standardized and fused using a stacking-based fusion approach, as defined in Equation (4), to construct the MCAFF representation $M_i(t)$.

4.3. CNN-Based Classification

The MCAFF representation $M(t) \in \mathbb{R}^{l \times z}$ serves as input for a CNN aimed at classifying conditions. The CNN architecture is structured to learn hierarchical feature representations through 3 sequential convolutional blocks, closing in a fully connected classifier. Table 1 describes the architecture used for this condition classification (Table 1).

Table 1. The CNN structure employed in the proposed study

Block	Layer Composition	Output Channels	Regularization
Block 1	Conv–BN–ReLU–MaxPool	32	Dropout (0.25)
Block 2	Conv–BN–ReLU–MaxPool	64	Dropout (0.25)
Block 3	Conv–BN–ReLU–MaxPool	128	Dropout (0.30)
Classifier	Flatten–FC–BN–ReLU–Dropout	256	Dropout (0.40)
Output Layer	Fully Connected	3 Classes	

Each convolutional block is comprised of several key components: a 3×3 convolution with a stride of 1 and padding of 1, followed by batch normalization to stabilize the learning process, a ReLU activation function to introduce non-linearity, and max pooling to reduce dimensionality and extract key features. To combat overfitting, dropout regularization is applied at multiple layers of the model. The classifier section includes a flatten layer that prepares the data for processing, followed by fully connected layers enhanced with batch normalization and ReLU activation. Dropout is again applied in this region to ensure robust training. Ultimately, the network's final output layer is a fully connected layer representing 3 distinct classes, each corresponding to different operating conditions.

The CNN model experienced training with a focus on minimizing the categorical cross-entropy loss function. This function plays a crucial role in assessing the disparity between the predicted probabilities generated by the model and the actual labels of the data. To enhance performance, the Adam optimizer was selected for its proficiency in managing sparse gradients and accelerating convergence.

5. Results and Discussion

This section discusses the condition classification of RBM systems and compares the results of the proposed method with those of two baseline methods. A strict experimental protocol was followed to ensure robustness and reliability, which included conducting 10 independent runs for a thorough evaluation. Figure 3 displays the multichannel feature fusion images corresponding to each speed condition of the robots.

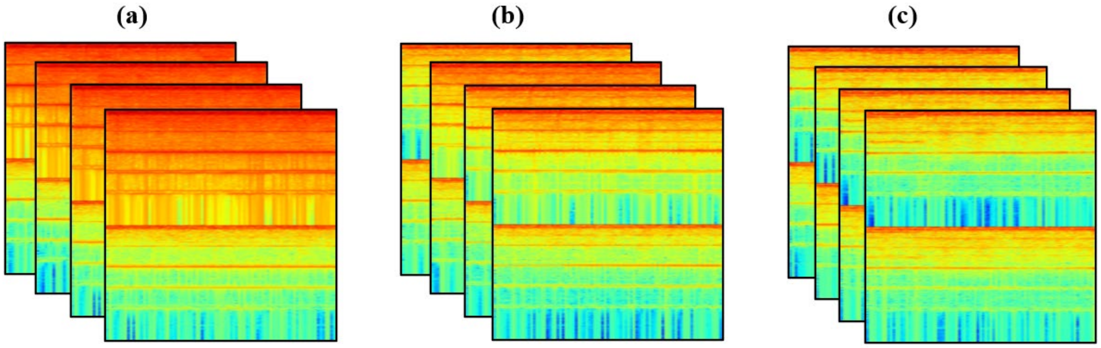

Figure 3: MCAFF corresponding to each speed condition of the robots: (a) 200 mm/s, (b) 500 mm/s, (c) 1000 mm/s.

5.1 Classification Performance of the Proposed MCAFF Method

The results were analyzed by averaging the outcomes across 10 trials, which effectively mitigated the effects of randomness and led to a more consistent and reliable assessment of the model's performance. When assessing model performance in situations with imbalanced classes, it is essential to utilize metrics like true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These metrics illustrate how well the model distinguishes between the classes. F1 scores ($F1-score = (TP)/[TP+1/2(FP+FN)]$) are computed to generate a balance between precision and recall, offering a unified measure that is especially beneficial for imbalanced data sets. The standard deviation (SD) values of the F1 scores reflect the variability observed across various evaluation attempts. The findings, including F1 scores and their corresponding SD values, are summarized in Table 2, facilitating a comparison of model performance in the context of condition classification.

Table 2. Condition based performance of the proposed method

Speed of the robot and label	Precision (SD)	Recall (SD)	Avg. F1-score per class (SD)
200 mm/s	0.9625 (0.0287)	0.9550 (0.0173)	0.9625 (0.0050)
500 mm/s	0.845 (0.0387)	0.9000 (0.0648)	0.8700 (0.0316)
1000 mm/s	0.8975 (0.0492)	0.8400 (0.0469)	0.8700 (0.0271)

Table 2 presents the precision, recall, and average F1-score for each robot speed condition: 200 mm/s, 500 mm/s, and 1000 mm/s. The data show that at 200 mm/s, the model performs well, achieving high precision (0.9625 ± 0.0287) and recall (0.9550 ± 0.0173), which results in a very stable F1-score of 0.9625 ± 0.0050. At 500 mm/s, there is slightly more variability, particularly in recall, which stands at 0.9000 ± 0.0648. This variability is also evident in the F1-score, which is 0.8700 ± 0.0316. For the speed of 1000 mm/s, precision is recorded at 0.8975 ± 0.0492, which is higher than recall (0.8400 ± 0.0469). The average F1-score remains solid at 0.8700 ± 0.0271, demonstrating good overall performance with moderate variability. These results indicate that the proposed method maintains robust performance across all robot operating conditions. Additionally, the standard deviation values provide confidence in the stability and reliability of the classification approach. Class-wise performance is illustrated in the confusion matrix shown in Figure 4.

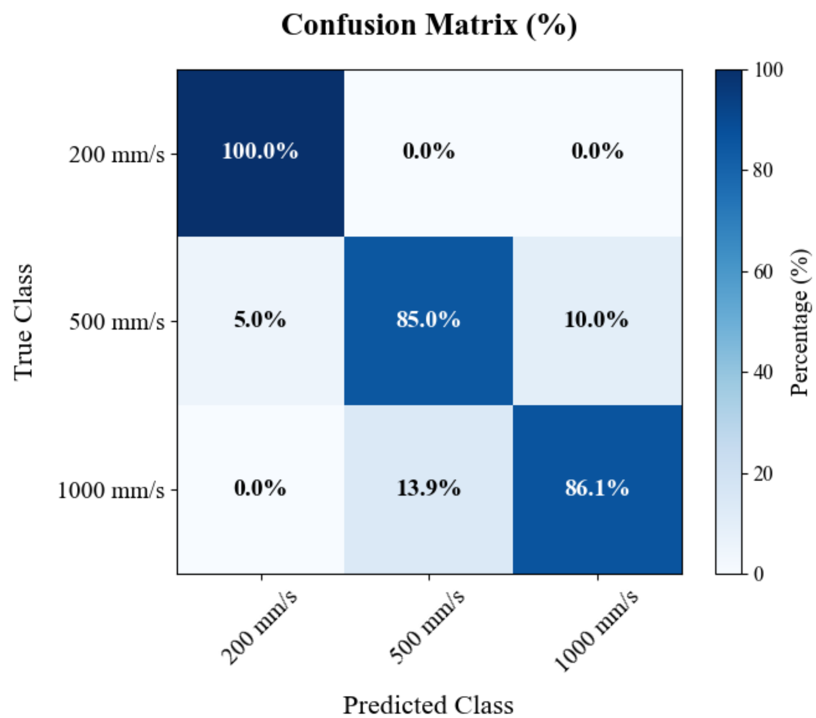

Figure 4. Confusion matrix for speed classification using test data.

The confusion matrix in Figure 4 illustrates the classification performance of the proposed MCAFF framework on the test set. The model successfully classified the majority of samples across all operating conditions. Minor misclassifications primarily occurred between end-to-end speed conditions (500 mm/s and 1000 mm/s), highlighting the advanced acoustic differences between these states.

5.2 Performance Comparison

To thoroughly evaluate the effectiveness of the proposed method, we conducted a comparison study with detailed analysis of both absolute and relative improvements. The results were compared against two baseline approaches: (i) single-sensor acoustic input and (ii) data-level fusion. The classification performance metrics are summarized in Table 3.

Table 3. Performance comparison of the proposed method

Method	**Precision (SD)**	**Recall (SD)**	**Avg. F1-score per class (SD)**
Single sensor	0.8375 (0.0263)	0.7558 (0.0143)	0.7425 (0.0109)
Data-level fusion	0.8766 (0.0301)	0.8183 (0.0179)	0.8133 (0.0888)
MCAFF (Proposed)	0.9016 (0.0388)	0.8983 (0.0430)	0.9008 (0.0212)

The proposed MCAFF method achieved an average F1-score of 0.9008, which is significantly higher than the 0.7425 score for the single-sensor approach and the 0.8133 score for data-level fusion. In terms of absolute improvement, MCAFF provides an approximately 15% increase over the single-sensor method and an approximately 10% increase over data-level fusion. The relative improvements further emphasize the effectiveness of the proposed approach: an approximately 21% improvement compared to the single-sensor method and approximately 10% improvement compared to data-level fusion. These results clearly demonstrate that the MCAFF framework significantly enhances classification performance by effectively leveraging multi-channel acoustic information.

5.3 Validation

The proposed MCAFF method achieves the highest classification performance while maintaining relatively low variability. Although the data-level fusion method shows higher variability, the MCAFF approach shows a strong balance between performance and stability. The consistency of the MCAFF method across repeated trials indicates its robustness to data variation. Moreover, the relatively low standard deviation compared to the performance gain suggests that the improvement is not merely due to random fluctuations but represents a genuine enhancement in model capability. Statistical significance was assessed using a two-sample t-test, confirming that the improvements achieved by the MCAFF method over both baseline approaches are statistically significant ($p < 0.001$).

These results highlight the significance of incorporating spatially distributed acoustic information through the MCAFF representation. The increased classification capability suggests that MCAFF effectively leverages diverse acoustic data, ultimately leading to superior performance in RBM applications. This indicates that the method not only advances the state of the art but also opens new possibilities for more accurate, robust classification in complex RBM environments.

5.4 Discussion

The performance of the MCAFF can be attributed to several crucial factors that enhance its ability to distinguish complex acoustic patterns. The implementation of multi-channel acoustic sensing provides a substantial advantage by capturing spatially diverse sound information that traditional single-sensor approaches cannot access. This spatial diversity contributes significantly to the framework’s ability to represent complex system dynamics more comprehensively. By leveraging multiple sensors positioned at different locations, the MCAFF framework can create a more holistic understanding of the acoustic environment, allowing it to differentiate between various sound signatures more effectively. The stacking-based feature fusion strategy employed in the MCAFF retains the unique characteristics of each channel while facilitating joint learning across channels. This methodology yields feature representations that are not only richer but also more discriminative, enabling the system to make subtle distinctions among different acoustic signals. The ability to combine information from multiple channels while preserving their distinct features improves classification and identification of variations in operating conditions, which is critical in many industrial applications. Additionally, CNN architecture plays a pivotal role in harnessing the structured nature of the MCAFF representation. By learning hierarchical spatial features, CNN is particularly adept at recognizing patterns and anomalies in acoustic data, even in the face of minor variations in conditions. This capacity to focus on different levels of abstraction from raw acoustic signals to higher-level features enables effective discrimination among operational states.

5.5 Limitations and Future Work

Despite the promising outcomes demonstrated by the MCAFF framework, several limitations need consideration. Firstly, the current investigation is based on a limited set of operating conditions, primarily focusing on velocity variations. This narrow scope may restrict the generalizability of the findings. Future research should explore a broader range of operating conditions, including more complex dynamic scenarios that better reflect real-world industrial settings. Secondly, as this is a pilot study, it used only two acoustic sensors. While this configuration has demonstrated positive results, increasing the number of sensors could further enhance the framework’s performance. More channels could provide additional layers of data, leading to even richer acoustic feature sets and improving the system’s capability to distinguish between closely related sound signatures. Moreover, the experiments were predominantly conducted in a controlled environment, which may not adequately replicate the complexities and noise present in actual industrial settings. Investigating the framework’s performance in noisy environments will be essential to assess its practicality and effectiveness in real-world applications.

Looking ahead, future work will aim to extend the MCAFF framework to encompass more complex fault scenarios, incorporating additional sensing modalities such as vibration and temperature sensors. This multi-modal approach is expected to enhance diagnostic accuracy by providing complementary data that can inform the understanding of equipment states. Furthermore, validating the framework in real-world industrial environments will be a crucial step in confirming its efficacy and versatility in practical applications.

6. Conclusion

This study introduces a MCAFF framework for classifying the operating conditions of a high-speed RBM using spatially distributed acoustic sensors. The framework integrates multi-channel time-domain acoustic signals and transforms them into a unified time-frequency representation, which is then processed by a CNN to capture complex patterns for accurate classification. Experimental results demonstrate that the proposed MCAFF approach significantly outperforms baseline methods. Specifically, it achieves an average F1-score of 0.9008 ± 0.0212, compared to

0.7425 ± 0.0109 for the single-sensor approach and 0.8133 ± 0.0888 for data-level fusion. Statistical validation confirms that these improvements are highly significant ($p < 0.001$), indicating superior classification performance and consistent behavior across repeated trials. The improved performance of the MCAFF framework can be attributed to the spatial diversity of multi-channel acoustic sensing, the stacking-based feature fusion strategy, and the hierarchical feature learning capability of CNNs. These components collectively enable effective discrimination of variations in acoustic signatures associated with different operating conditions. The proposed MCAFF framework provides a robust and effective solution for RBM condition monitoring, demonstrating strong potential for applications in predictive maintenance. Future work will focus on extending the framework to more complex scenarios and validating its performance in real-world industrial environments.

.

Acknowledgements
The authors would like to thank Tennessee Technological University for supporting this research.

References
Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H. and Adda, M., On predictive maintenance in Industry 4.0: Overview, models, and challenges, *Applied Sciences*, vol. 12, no. 16, pp. 8081, 2022, https://doi.org/10.3390/app12168081.

Al Mamun, A., Bappy, M.M., Senanayaka, A.M., Li, J., Jiang, Z., Tian, Z., Fuller, S., Falls, T.C., Bian, L. and Tian, W., Multi-channel sensor fusion for real-time bearing fault diagnosis by frequency-domain multilinear principal component analysis, *International Journal of Advanced Manufacturing Technology*, vol. 124, pp. 1321–1334, 2023, https://doi.org/10.1007/s00170-022-10525-4.

Gong, C.-S., Su, C.-H. S., Chao, K.-W., Lin, C.-Y. and Chen, Y.-H., Acoustic-based fault detection for robotic arms, *IEEE Journal of Selected Areas in Sensors*, vol. 3, pp. 1-22, 2025, https://doi.org/10.1109/JSAS.2025.3638975.

Hassan, I. U., Panduru, K. and Walsh, J., An in-depth study of vibration sensors for condition monitoring, *Sensors*, vol. 24, no. 3, 740, 2024, https://doi.org/10.3390/s24030740.

Jardine, A. K. S., Lin, D. and Banjevic, D., A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483-1510, 2006, https://doi.org/10.1016/j.ymssp.2005.09.012.

Kullu, O. and Cinar, E., A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults, *Machines*, vol. 10, no. 11, 1105, 2022, https://doi.org/10.3390/machines10111105.

Kumar, S., Goyal, D., Dang, R. K., Dhami, S. S. and Pabla, B., Condition based maintenance of bearings and gears for fault detection—A review, *Materials Today: Proceedings*, vol. 5, no. 2, pp. 6128-6137, 2018, https://doi.org/10.1016/j.matpr.2017.12.219.

Lee, N., Senanayaka, A., Storey, J., Murphy, M., Hwang, J., Lim, H., Cai, B., Netchaeve, A. and Stone, T., Enhancing defect detection accuracy in wire-arc additive manufactured samples through fusion of multi-directional scanning with phased array ultrasound testing (PAUT), *Proceedings of the ASME 2025 International Mechanical Engineering Congress and Exposition (IMECE2025)*, Memphis, TN, USA, November 16–20, 2025, IMECE2025-165020, https://doi.org/10.1115/IMECE2025-165020.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N. and Nandi, A. K., Applications of machine learning to machine fault diagnosis: A review and roadmap, *Mechanical Systems and Signal Processing*, vol. 138, 106587, 2020, https://doi.org/10.1016/j.ymssp.2019.106587.

Li, Y., Daho, M. E. H., Conze, P.-H., Zeghlache, R., Le Boité, H., Tadayoni, R., Cochener, B., Lamard, M. and Quellec, G., A review of deep learning-based information fusion techniques for multimodal medical image classification, *Computers in Biology and Medicine*, vol. 177, 108635, 2024, https://doi.org/10.1016/j.compbiomed.2024.108635.

Makulavičius, M., Petkevičius, S., Rožėnė, J., Dzedzickis, A. and Bučinskas, V., Industrial robots in mechanical machining: Perspectives and limitations, *Robotics*, vol. 12, no. 6, 160, 2023, https://doi.org/10.3390/robotics12060160.

Mobley, R. K., *An introduction to predictive maintenance*, 2nd edition, Butterworth-Heinemann, 2002, https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3.

Ntalampiras, S., One-shot learning for acoustic diagnosis of industrial machines, *Expert Systems with Applications*, vol. 178, 114984, 2021, https://doi.org/10.1016/j.eswa.2021.114984.

Peng, Y., Dong, M. and Zuo, M. J., Current status of machine prognostics in condition-based maintenance: A review, *International Journal of Advanced Manufacturing Technology*, vol. 50, no. 1, pp. 297-313, 2010, https://doi.org/10.1007/s00170-009-2482-0.

Piankitrungreang, P., Chaiprabha, K., Chungsangsatiporn, W., Ratanasumawong, C., Chancharoen, P. and Chancharoen, R., Acoustic-based machine main state monitoring for high-speed CNC drilling, *Machines*, vol. 13, no. 5, p. 372, 2025, https://doi.org/10.3390/machines13050372.

Qiao, G. and Weiss, B. A., Accuracy degradation analysis for industrial robot systems, *Proceedings of the ASME 2017 12th International Manufacturing Science and Engineering Conference (MSEC2017)*, 2017, https://doi.org/10.1115/MSEC2017-2782.

Qurthobi, A., Maskeliūnas, R. and Damaševičius, R., Detection of mechanical failures in industrial machines using acoustic signals: A systematic review, *Sensors*, vol. 22, no. 10, 3888, 2022, https://doi.org/10.3390/s22103888.

Senanayaka, A., Lee, P., Lee, N., Dickerson, C., Netchaev, A. and Mun, S., Enhancing the accuracy of machinery fault diagnosis through fault source isolation of complex mixture of industrial sound signals, *International Journal of Advanced Manufacturing Technology*, vol. 133, pp. 5627-5642, 2024, https://doi.org/10.1007/s00170-024-14080-y.

Senanayaka Mudiyanselage, A., Liu, Q., Lee, N., Mun, S., Amirlatifi, A., Jabour, J., Arnold, T. and Seale, M., Frequency domain tensor-based 1D-convolutional neural network and multilinear principal component analysis for machinery fault detection, *Proceedings of the Annual Conference of the PHM Society*, vol. 16, no. 1, 2024, https://doi.org/10.36001/phmconf.2024.v16i1.3871.

Soori, M., Dastres, R., Arezoo, B. and Karimi Ghaleh Jough, F., Intelligent robotic systems in Industry 4.0: A review, *Journal of Advanced Manufacturing Science and Technology*, vol. 4, no. 3, 2024007, 2024, https://doi.org/10.51393/j.jamst.2024007.

Tama, B. A., Vania, M., Lee, S. and Lim, S., Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals, *Artificial Intelligence Review*, vol. 56, no. 5, pp. 4667-4709, 2023, https://doi.org/10.1007/s10462-022-10293-3.

Tang, Q., Liang, J. and Zhu, F.., A comparative review on multi-modal sensors fusion based on deep learning, *Signal Processing*, vol. 213, 109165, 2023, https://doi.org/10.1016/j.sigpro.2023.109165.

Tsanousa, A., Bektsis, E., Kyriakopoulos, C., González, A. G., Leturiondo, U., Gialampoukidis, I., Karakostas, A., Vrochidis, S. and Kompatsiaris, I., A review of multisensor data fusion solutions in smart manufacturing: Systems and trends, *Sensors*, vol. 22, no. 5, 1734, 2022, https://doi.org/10.3390/s22051734.

Uhlmann, E., Polte, J. and Geisert, C., Condition monitoring concept for industrial robots, *Proceedings of the 17th IMEKO TC10 Conference "Global Trends in Testing, Diagnostics & Inspection for 2030"*, October 20–22, 2020.

Wang, D., Li, Y., Jia, L., Song, Y. and Liu, Y., Novel three-stage feature fusion method of multimodal data for bearing fault diagnosis, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-10, 2021, https://doi.org/10.1109/TIM.2021.3071232.

Wang, S., Jiang, L., Meng, J., Xie, Y., and Ding, H., Training for smart manufacturing using a mobile robot-based production line, *Frontiers of Mechanical Engineering*, vol. 16, no. 2, pp. 249-270, 2021, https://doi.org/10.1007/s11465-020-0625-z.

Xiao, X., Zhang, Y., Li, X., Wang, H. and Chen, J., Rotating machinery fault diagnosis method based on multi-level information fusion, *Information Fusion*, vol. 113, 2025, https://doi.org/10.1016/j.inffus.2024.102621.

Biographies

Md Omar Al Javed is a Graduate Research Assistant in the Department of Mechanical and Nuclear Engineering at Tennessee Technological University, Cookeville, Tennessee, USA. His research focuses on smart manufacturing, predictive maintenance, industrial robotics, and data-driven decision-making. He received his BS in Industrial and Production Engineering from the Military Institute of Science and Technology (MIST), Bangladesh. His background includes manufacturing process optimization, design of experiments, and machining process analysis. He has authored and co-authored several journal and conference publications in manufacturing and industrial engineering.

Ayantha Senanayaka is an Assistant Professor in the Department of Manufacturing and Engineering Technology at Tennessee Technological University. His research focuses on smart manufacturing, predictive maintenance, robotics, and non-contact sensing systems. He specializes in integrating machine learning with advanced sensing technologies for improving manufacturing system reliability and efficiency. His work emphasizes scalable, cost-effective solutions for industrial applications, particularly in high-speed robotic environments. He has published in leading journals and conferences in manufacturing and industrial engineering and actively contributes to interdisciplinary research in intelligent manufacturing systems.