Received: 02 Nov. 2025; Revised: 12 Jan. 2026; Accepted: 21 Jan. 2026
https://doi.org/10.22306/al.v13i2.780

Multi-class fault classification in conveyor systems using machine learning: enhancing reliability in production logistics

Hassan Hijry
Department of Industrial Engineering, University of Tabuk, Tabuk 47512, Saudi Arabia,
hhegri@ut.edu.sa

Keywords: industrial conveyor systems, multi-fault detection, predictive maintenance, production automation.

Abstract: Industrial conveyor systems are vital to modern manufacturing and logistics operations, where unexpected failures can cause substantial operational disruptions in material flows and economic losses. Traditional maintenance strategies often fall short in addressing the complex and interrelated failure patterns present in contemporary conveyor systems. This study presents a machine learning-based framework for developing multi-class classification models, designed to enable reliable deployment for real-time fault diagnosis in logistics environments. Using real-world operational data with sensor-derived features, multiple machine learning models are trained and evaluated to classify key types of conveyor faults. Experimental results demonstrate that ensemble methods achieved the highest performance, attaining an accuracy of 92.56% and significantly outperforming linear and instance-based approaches. This research aims to advance predictive maintenance by introducing a unified framework for the development of a multi-class classification model, enabling the identification of the most effective model for deployment in logistics settings. The main contribution of this work lies in the integration of machine learning techniques into logistics systems for predictive maintenance, offering an advanced, scalable solution that can be deployed to improve system reliability, reduce downtime, and enhance operational efficiency.

1 Introduction

Industrial conveyor systems represent a cornerstone of modern manufacturing, logistics, and material transport across diverse sectors, including mining, agriculture, food processing, automotive, and distribution centers. Conveyors are commonly used to move materials efficiently, helping ensure smooth operations and consistent productivity across various environments. The widespread application of industrial conveyors across various sectors underscores their critical role in logistics and material flow management in modern industrial operations. They not only improve the efficiency of material transport but also help reduce labor costs, enhance worker safety, and increase overall productivity. As industries continue to evolve and expand, the role of conveyors is becoming even more vital, with innovations such as automation, artificial intelligence, and robotics driving further improvements in performance, capabilities, and logistics optimization. Figure 1 illustrates the extensive range of applications for industrial conveyors across these sectors, highlighting their integral role in streamlining processes and boosting productivity in modern industrial operations.

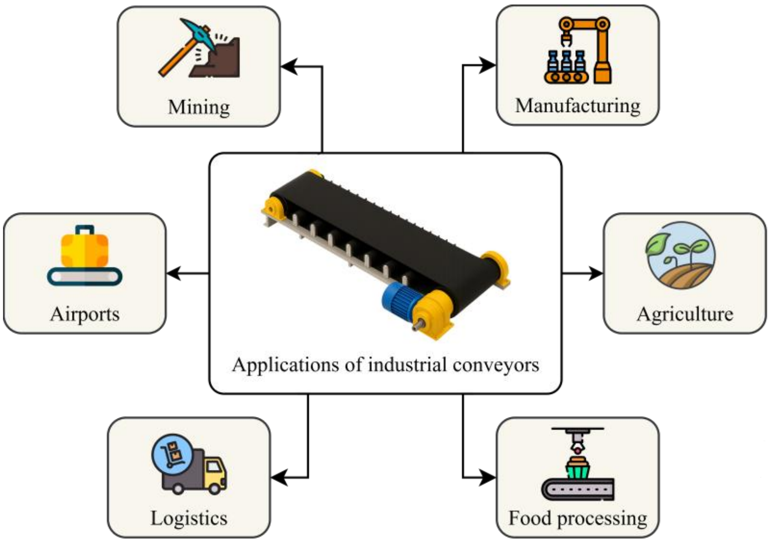


Figure 1 Common applications of industrial conveyors

Conveyor system failures can lead to significant financial losses in industrial operations by disrupting material flows, increasing downtime, and driving up maintenance costs. These unplanned downtimes resulting from component failures can halt entire production lines, leading to immediate revenue losses and cascading effects throughout the logistics chain. Beyond direct operational costs, unexpected failures often necessitate emergency repairs, premium pricing for expedited parts procurement, and extended maintenance personnel overtime. Safety concerns further compound these issues, as sudden mechanical failures can pose risks to personnel working in proximity to the equipment. The fundamental design of a belt conveyor system, as depicted in Figure 2, consists of several interconnected mechanical components, including drive motors, gearboxes, conveyor belts, rollers, pulleys, and bearings. Each component plays a vital role in the system's overall functionality, and the failure of any element can result in a complete system shutdown, leading to significant operational disruptions and economic losses.

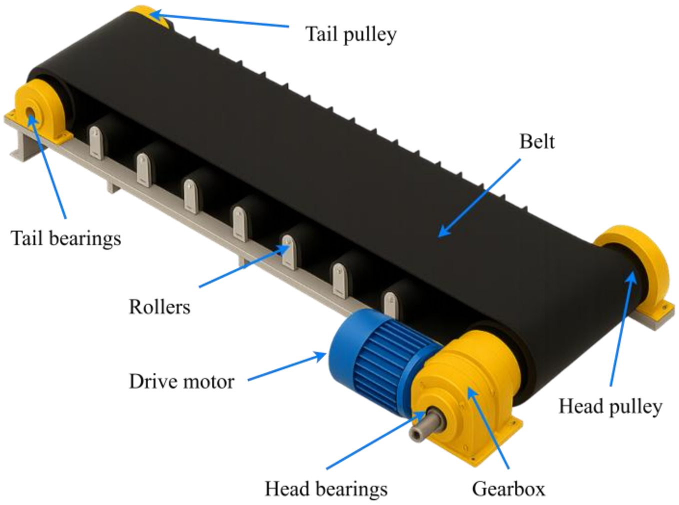


Figure 2 Key components of a belt conveyor: drive motor, gearbox, conveyor belt, rollers, pulleys, and bearings

Traditional maintenance approaches, predominantly based on periodic maintenance or reactive strategies following equipment failure, have proven inadequate for addressing the complex failure patterns exhibited by modern conveyor systems [1,2]. Periodic maintenance often results in unnecessary component replacements and increased operational costs, while reactive maintenance leads to unexpected downtime and potential safety hazards, disrupting material flows and logistics operations. The limitations of these conventional approaches have driven the industrial sector toward predictive maintenance strategies that leverage advanced monitoring technologies and data analytics to anticipate failures before they occur. The evolution of conveyor fault detection has progressed through several distinct phases, each characterized by technological advancements and improved diagnostic capabilities. Early detection methods relied primarily on manual inspections and basic vibration monitoring, providing limited insight into component health and often failing to detect incipient failures [1]. The introduction of vibration and acoustic signal analysis marked a significant advancement, enabling the detection of bearing and rotating component defects through frequency domain analysis [3,4].

Recent technological advancements have enhanced diagnostic capabilities, incorporating tools like acoustic monitoring, thermal imaging, and computer vision techniques. Acoustic signal analysis has emerged as a particularly effective approach for identifying bearing and idler roller defects, with researchers demonstrating the successful application of signal processing techniques such as Fast Fourier Transform (FFT) and Mel Frequency Cepstral Coefficient (MFCC) extraction for fault classification [5-7]. The integration of autonomous inspection systems has further enhanced diagnostic capabilities, with mobile robots equipped with acoustic sensors proving effective for continuous monitoring in challenging industrial environments [5,6]. The integration of machine learning and artificial intelligence has transformed conveyor fault detection from reactive monitoring to proactive predictive maintenance systems. Modern approaches utilize advanced techniques such as deep learning, anomaly detection, and reinforcement learning to analyze complex sensor data, uncover meaningful patterns, and predict component failures with high accuracy. Deep learning techniques have shown particular promise for visual inspection tasks, with Convolutional Neural Networks (CNNs) achieving remarkable performance in detecting belt damage and component defects from video feeds and thermal images [8-10]. Edge computing implementations have further enhanced the practical applicability of machine learning approaches, enabling real-time fault detection with reduced computational overhead and improved response times [11].

While these technological advancements have significantly enhanced diagnostic capabilities and demonstrated impressive performance in controlled research environments, their translation into comprehensive industrial solutions remains challenging. The promising results achieved through acoustic monitoring, thermal imaging, and machine learning approaches have primarily been validated under specific operational conditions and targeted fault scenarios. However, the complexity of real-world industrial conveyor operations characterized by simultaneous multi-component interactions, varying operational loads, and diverse environmental conditions presents significant challenges for logistics operations that extend beyond the scope of current research focuses. The convergence of these technological limitations with evolving industrial requirements creates a compelling need for more sophisticated diagnostic solutions. Modern industrial facilities increasingly demand predictive maintenance systems that can provide actionable intelligence about specific fault conditions rather than generic anomaly alerts. The economic implications of unplanned downtime, which often exceed thousands of dollars per hour in manufacturing environments, necessitate diagnostic systems capable of distinguishing between fault types to enable targeted maintenance interventions. Moreover, the integration of Industry 4.0 principles requires condition monitoring systems that can seamlessly interface with existing industrial Internet of Things (IoT) infrastructures while providing granular diagnostic information that supports data-driven maintenance scheduling and resource allocation decisions.

This research addresses these critical needs through the development of a framework capable of training and selecting a multi-class fault classification model specifically designed for conveyor system diagnostics. The study establishes two primary research objectives: first, to develop and validate a machine learning-based diagnostic system capable of simultaneously identifying six distinct conveyor fault categories (ball bearing, central shaft, pulley, drive motor, idler roller, and belt slippage); and second, to conduct a comprehensive comparative evaluation of multiple classification algorithms under realistic operational conditions, including detailed feature importance analysis that identifies the most discriminative features for each fault type. The best model identified by the proposed framework is suitable for deployment in logistics environments, where it can provide real-time fault classification and support predictive maintenance decision-making processes. The modular design approach adopted ensures that the proposed model development framework can be adapted for deployment across diverse industrial conveyor installations through appropriate sensor integration and system calibration procedures.

The rest of this paper is organized as follows: Section 2 provides a comprehensive review of related work in conveyor fault detection, examining existing approaches and identifying research gaps that motivate this study. Section 3 details the methodology, describes the dataset used for model development and evaluation, including fault categorization and statistical characteristics, encompassing data preprocessing techniques, feature extraction methods, model training procedures, and hyperparameter tuning strategies. Section 4 discusses experimental results, including comparative performance, feature importance analysis, and practical implications for deployment in logistics. Finally, Section 5 concludes the paper with a summary of key findings, contributions to the field, and recommendations for future research directions.

2 Literature review

Conveyor fault detection has become an increasingly active area of research, with notable progress in both theoretical development and industrial applications. Initially, fault detection relied on manual inspections; however, recent advancements in technology have introduced machine learning, computer vision, distributed sensing, and hybrid approaches. These innovations are aimed at improving system reliability, enabling predictive maintenance, and enhancing cost efficiency in conveyor operations. Conveyor systems are prone to various faults, which can significantly impact their performance and safety. Common faults include belt misalignment or deviation, belt slippage, belt tearing, as well as failures in idlers, bearings, pulleys, and gears. These issues typically arise due to factors such as uneven loading, adverse environmental conditions, inadequate maintenance, or mechanical wear. These faults can result in costly downtime and pose safety risks, making their detection crucial. This section systematically examines the literature on conveyor fault detection and the limitations of the existing approaches.

Traditional approaches, such as periodic manual inspection and monitoring of sensor outputs, offer limited fault coverage and may not catch early-stage defects [1]. Vibration and acoustic signal-based detection have emerged as a standard, particularly for idler and bearing issues. Techniques such as FFT and MFCC extraction form the basis for signal analysis, with machine learning playing a growing role in classification and anomaly detection [3]. Skoczylas et al. developed an autonomous legged robot that uses acoustic signals to detect faulty idler rollers in underground conveyor systems. Their approach combines cyclostationary signal analysis with Principal Components Analysis (PCA) to enable robust fault detection [5]. Wodecki et al. refined the acoustic diagnostic pipeline by incorporating cyclic spectral coherence, local mode decomposition, and envelope analysis. The proposed method improved the robot's ability to distinguish real bearing faults from environmental or structural noise artifacts [6]. Wijaya et al. proposed an automated conveyor fault detection system using distributed acoustic sensing via optical fibers, enabling long-range, real-time vibration monitoring. Their method identifies abnormal roller behavior by analyzing acoustic signal patterns and applies automated decision rules to reduce inspection delays in mining environments [2]. Liu et al. propose a fault detection system for belt conveyor idlers using acoustic signals processed through 13 MFCCs. A Gradient Boosted Decision Tree (GBDT) model is trained on these features to classify the idler condition, achieving 94.53% accuracy and up to 99.7% recall on test data [7]. Milovancevic explored diagnoses of belt conveyor idler faults using vibration signals. The study employs Short-Time Fourier Transform (STFT) and CNN for effective fault classification under various operational conditions, enhancing maintenance reliability [4].

Thermal and acoustic imaging combined with computer vision and machine learning techniques have been widely applied for early, non-invasive conveyor fault detection. Bortnowski et al. explored the use of an acoustic camera combining a video feed with a microphone array for spatial mapping of noise sources on a belt conveyor in a controlled laboratory setup. The authors recorded and visualized sound pressure level maps to locate and characterize noise from three principal elements: the electric motor, idler roller bearings, and tail pulley misalignment [12]. Fedorko et al. proposed a methodology for identifying noise sources in continuous transport systems using an acoustic camera. The proposed approach combines beamforming and spectral analysis to localize and classify mechanical noise from components like idlers and motors, supporting condition monitoring and preventive maintenance [13]. Thermal imaging combined with computer vision and machine learning techniques has been widely applied for early, non-invasive conveyor fault detection by analyzing temperature anomalies. Yang et al. introduced a motor-driven inspection robot that uses infrared thermography to monitor critical mechanical components of belt conveyors. By processing thermal images of parts like motors, pulleys, and rollers, the robot automatically detects abnormalities and issues failure warnings, overcoming the limitations of traditional inspection methods [14]. Siami et al. proposed a binary classification method using CNN for identifying overheated belt conveyor idlers using thermal images. The proposed method achieved a precision of 0.9740 and an F1 score of 0.9782, significantly improving previous results [15]. Zhan et al. proposed a deeply lightweight target detection network based on the Yolov4 for foreign object detection on conveyor belts. The proposed system offers superior speed and accuracy, enabling real-time alerts that support timely interventions and minimize system downtime [16]. Xiuyu et al. proposed a method for diagnosing faults in conveyor rollers using thermal infrared imaging. The approach utilizes the YOLOv4 vision method to accurately locate rollers. This method effectively distinguishes between normal and faulty rollers based on temperature changes observed during operation, achieving a recognition accuracy of 93.8% [17]. Siami et al. proposed an image processing pipeline using U-Net-based CNNs enhanced with thermal image augmentation for automatic semantic segmentation of thermal defects in belt conveyor idlers. The proposed approach improved detection accuracy for thermal anomalies, achieving a mean pixel accuracy of 99.9%, supporting more reliable, automated fault monitoring in challenging industrial environments [18].

In recent years, machine learning and deep learning techniques have gained significant attention for conveyor fault detection, offering advanced capabilities for real-time and accurate monitoring. Various studies have demonstrated the effectiveness of machine learning and deep learning in detecting conveyor faults, enabling timely fault identification and predictive maintenance. Li et al. developed a hybrid fault diagnosis model for belt conveyors, integrating support vector machine (SVM), principal component analysis, and grey wolf optimization. The methodology addresses the limitations of standard optimization methods, achieving a fault classification accuracy of 97.22% using monitoring data from underground mine conveyors [19]. Zhang et al. proposed a lightweight deep learning method for detecting damage in mining conveyor belts by integrating MobileNet with YOLOv4. It achieves 93.22% accuracy and 70.26 FPS speed, outperforming the original Yolov4 and demonstrating strong generalization for visual monitoring [8]. Liu et al. developed a deep learning-based method for detecting damage in mining conveyor belts using on-site monitoring video. It combines temporal and spatial features, employing an improved attention mechanism and Temporal Convolutional Networks, achieving over 20% higher accuracy than traditional methods [9]. Dwivedi et al. developed a deep learning framework for real-time damage detection in long conveyor belts, utilizing edge devices for immediate results. It effectively identifies damage sizes from 1 cm to 100 cm, achieving an 85% mean average precision in tunnel construction sites [10]. Soares et al. highlighted the use of predictive techniques like vibration analysis and machine learning to address reliability issues in bulk transportation systems, specifically belt conveyors. By combining Wavelet Packet Decomposition for feature extraction and GBDT for fault classification, the system achieved high accuracy (100% for bearing faults and 97.5% for surface wear), proving effective for diagnosing issues in conveyor idlers [20]. Gunckel et al. proposed a flexible machine learning workflow for developing failure forecasting systems for mining conveyor belts in Chile. The approach involved integration between various components, including the distributed control system, a digital twin, and an operational logbook. The proposed approach achieved precision and recall above 0.83, reducing dependence on maintenance data [11]. Liu et al. proposed an intelligent fault diagnosis method for belt conveyor rollers using a polar KNN algorithm with audio features. By extracting and analyzing audio signals, their method enhances fault classification accuracy, outperforming traditional KNN and similar classifiers for detecting roller defects [21].

Although significant advancements have been made in conveyor fault detection through the application of acoustic sensing, thermal imaging, computer vision, and machine learning techniques, the majority of existing studies concentrate on the detection of a single fault type or component. Commonly investigated faults include idler roller failure, bearing defects, or belt surface damage, each typically addressed in isolation. While these focused approaches have demonstrated high accuracy within controlled settings, they often lack scalability and practical relevance in logistics, where multiple faults can occur simultaneously, disrupting material flows. Despite the growing interest in predictive maintenance, relatively few studies have proposed diagnostic frameworks capable of identifying and distinguishing multiple fault types within a unified model. There remains a clear gap in the literature for multi-class classification techniques that can generalize across fault categories and support multi-fault diagnosis. To address this shortcoming, the present study introduces a machine learning-based classifier designed to detect and differentiate six critical types of conveyor faults: ball bearing, central shaft, pulley, drive motor, idler roller, and belt slippage. By targeting multiple components within a single diagnostic model, this research aims to enhance fault coverage, improve practical applicability, and contribute to the development of more robust condition monitoring systems for logistic conveyor operations.

3 Methodology

The development of an accurate and reliable machine learning model for industrial conveyor fault detection requires a structured, transparent, and replicable pipeline. This research follows a methodical approach comprising three main stages: data preprocessing, model training and hyperparameter tuning, and finally model evaluation and Selection. The preprocessing stage ensures that the data is preprocessed and is suitable for model input. Model training with hyperparameter tuning ensures that each model is finetuned for the fault classification task, and performance evaluation helps identify the best model for the classification task. The following text describes each step in detail.

3.1 Dataset description

This study employs an open-source operational conveyor fault dataset, publicly available on Kaggle, a well-known and reputable source of machine learning datasets. The dataset comprises 1,209 operational records collected from a conveyor system operating under industrial conditions. A distinguishing characteristic of this dataset lies in its multi-fault coverage, which includes six real-world conveyor faults: Ball Bearing Fault, Central Shaft Fault, Pulley Fault, Drive Motor Fault, Idler Roller Fault, and Belt Slippage. This contrasts with many publicly available industrial datasets, which often focus on detecting a single fault type. The multi-fault nature of the dataset significantly enhances its applicability to real-world industrial diagnostics, where simultaneous or interacting faults are common. Each record in the dataset encapsulates a detailed snapshot of system performance, capturing five key sensor-derived features: rotational speed, load, temperature, vibration, and electric current. A summary of feature types and their respective measurement units is provided in Table 1.

Table 1 Dataset features with measurement units

Features	Data types	Unit
Speed	Integer	rpm (rotations per minute)
Load	Integer	Kg (kilogram)
Temperature	Integer	°C (degree Celsius)
Vibration	Float	m/s^2 (meters per second squared)
Current (A)	Float	A (ampere)

Some random samples from the dataset for each fault class are presented in Table 2, showing how different operational parameters correspond to various faults. Since the dataset is based on variable load conditions, this operational variability makes it highly suitable for machine learning tasks such as predictive maintenance, fault classification, anomaly detection, and multi-class modelling. Also, its open-source availability on Kaggle promotes transparency, reproducibility, and benchmarking, making it a valuable resource for academia and industry alike in the pursuit of reliable, efficient conveyor diagnostics. The next section presents the methodology used to develop a multiclass fault classifier using this dataset.

Table 2 Random data samples with associated faults

Speed	Load	Temperature	Vibration	Current	Fault
116	490	43	0.82	3.17	Ball bearing
123	503	43	0.88	3.54	Central shaft
123	507	39	0.86	3.57	Pulley
122	535	44	1.06	3.56	Drive motor
118	509	40	0.8	3.48	Idler roller
116	473	40	0.77	3.23	Belt slippage

3.2 Data preprocessing

Data preprocessing is the crucial step of preparing raw input data before feeding it into machine learning algorithms. This step encompasses essential tasks such as handling missing values, normalizing feature scales, and partitioning the dataset into appropriate subsets, among other transformations. This process ensures data quality, consistency, and compatibility with machine learning models, directly impacting model performance and reliability. Effective preprocessing can significantly improve model accuracy, while poor preprocessing often leads to suboptimal results regardless of the algorithm used. For this study, since the dataset contained no missing values, the preprocessing steps were focused on feature scaling to normalize the input variables and partitioning the data into training, validation, and test sets. Feature scaling is an important step, especially when dealing with heterogeneous sensor data, like the one used in this study. To ensure optimal model performance and convergence, all numerical features were normalized using Min-Max scaling. This transformation maps each feature to a fixed range between 0 and 1, preventing features with larger numerical ranges from dominating the learning process. The formula for the Min-Max scaling transformation is presented in the following Equation (1).

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where $\boldsymbol{X}$ represents the original feature values, $\boldsymbol{X_{min}}$ and $\boldsymbol{X_{max}}$ are the minimum and maximum values of each feature, respectively. This scaling method was chosen over standardization (z-score normalization) as it preserves the original distribution shape while ensuring all features contribute equally to distance-based calculations commonly used in machine learning algorithms.

Another important preprocessing step was partitioning the dataset into subsets. Initially, the dataset was divided into training and test sets using an 80:20 ratio, with stratified sampling applied to maintain the original class distribution in both subsets. The test set was reserved exclusively for final model evaluation and remained untouched throughout the model development process. The training portion (80% of the original data) served as the foundation for model training and hyperparameter optimization. This training set was further subdivided during the model training phase using 5-fold cross-validation, as detailed in the subsequent model training section. This approach ensures that model performance estimates are derived from multiple independent validation sets while maintaining a held-out test set for unbiased final evaluation. The next steps in the developed pipeline are model training and evaluation.

3.3 Model training and hyperparameter tuning

To identify and select the best algorithm for conveyor fault detection, six distinct machine learning algorithms were selected to provide a comprehensive comparison of classification approaches, ranging from linear models to ensemble methods. The selected algorithms encompass different learning paradigms: Logistic Regression as a linear probabilistic classifier, Decision Tree as a rule-based approach, Random Forest as a bagging ensemble method, k-Nearest Neighbors as an instance-based learner, Support Vector Machine as a margin-based classifier, and XGBoost as a gradient boosting ensemble technique. This diverse selection ensures robust evaluation across various algorithmic approaches and provides insights into the most suitable modelling strategy for the given classification problem. To ensure each model achieves the best performance, hyperparameter optimization was conducted using an exhaustive grid search approach combined with 5-fold cross-validation to ensure robust parameter selection and prevent overfitting to specific data splits. The grid search methodology systematically evaluates all possible combinations of hyperparameters within predefined search spaces, providing comprehensive exploration of the provided parameter combinations for each algorithm. The 5-fold cross-validation procedure partitioned the training dataset into five equal stratified folds, maintaining the original class distribution within each fold.

Table 3 Hyperparameters and search space of grid search

Classifier	**Parameters and Search Space**
Logistic Regression	C: [0.1, 1, 10] solver: liblinear
Decision Tree	criterion: [gini, entropy] min samples leaf: [1, 2, 4] max depth: [None, 5, 10, 15, 20] min samples split: [2, 5, 10]
Random Forest	max depth: [None, 10, 20] n estimators: [50, 100, 200]
k-Nearest Neighbors	n neighbors: [3, 5, 7] weights: [uniform, distance]
Support Vector Machine	Kernel: [linear, rbf] C: [0.1, 1, 10]
XGBoost	learning rate: [0.01, 0.1, 0.2] max depth: [3, 6, 9] n estimators: [50, 100, 150]

For each hyperparameter combination, the model was trained on four folds and validated on the remaining fold, with this process repeated five times to ensure each fold served as the validation set exactly once. The final performance metric for each hyperparameter configuration was computed as the mean validation score across all five folds, providing a robust estimate of model performance while minimizing variance due to specific train-validation splits. Table 3 presents the defined search space of hyperparameters for various models.

Model selection was based on cross-validation accuracy, with the hyperparameter configuration yielding the highest mean validation score selected as optimal for each algorithm. The computational complexity of this approach resulted in training multiple model instances: for each algorithm, the total number of models trained equaled the Cartesian product of hyperparameter options multiplied by the number of cross-validation folds. This exhaustive evaluation ensures optimal hyperparameter selection while providing statistically robust performance estimates through repeated validation on different data subsets. Following hyperparameter optimization, final models were retrained using the complete training dataset with optimal parameters, preparing them for evaluation on the held-out test set. The following subsection describes the model evaluation process and metrics used to assess and select the best model.

3.4 Model evaluation

Following hyperparameter optimization through grid search and cross-validation, the best-performing configuration for each algorithm was selected based on cross-validation accuracy. These optimized models were subsequently evaluated on the previously unseen (held out) test set to assess their generalization capability and provide unbiased performance estimates. The test set, comprising 20% of the original dataset, remained completely isolated during the model development phase to ensure evaluation of model performance on new, unseen data. The best model was selected based on various evaluation metrics, including accuracy, Matthew's Correlation Coefficient (MCC), precision, recall, and F1 score. These metrics collectively capture various aspects of classification performance, including overall correctness, class-specific performance, and balanced evaluation measures. The following text briefly describes each evaluation metric:

- **Accuracy:** Accuracy represents the proportion of correctly classified instances across all classes and serves as the most intuitive measure of overall model performance. Equation (2) shows the formula for calculating accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

where TPs (True Positives), TNs (True Negatives), FPs (False Positives), and FNs (False Negatives) represent the counts of four possible prediction outcomes in the confusion matrix. The confusion matrix organizes these counts in a total classes × total classes table (6 x 6 in our case), with rows representing actual classes and columns representing predicted classes, providing a comprehensive view of classification performance. For any specific class i, the metrics are calculated using a one-vs-all approach: TPs_i are instances of class i correctly predicted as class i, FPs_i are instances of other classes incorrectly predicted as class i, FNs_i are instances of class i incorrectly predicted as other classes, and TNs_i are instances of other classes correctly predicted as not being class i. This decomposition allows computation of class-specific performance metrics, which are then aggregated using macro-averaging to provide overall model performance measures that treat all classes equally, regardless of their frequency in the dataset. While accuracy provides a straightforward interpretation of model performance, it can be misleading in cases of class imbalance where high accuracy might be achieved by simply predicting the majority class.

- **Matthew's Correlation Coefficient (MCC):** Although accuracy is commonly used to measure classification performance, it is asymmetrical and can be affected by class imbalance problems. In contrast, the MCC score is a more reliable metric for evaluating classification performance. Equation (3) shows the formula to calculate the MCC score.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{3}$$

MCC ranges from -1 to +1, where +1 indicates perfect prediction, 0 represents random prediction, and -1 indicates completely incorrect prediction. Unlike accuracy, MCC provides a balanced measure that accounts for all four confusion matrix categories and remains informative even with unbalanced datasets.

- **Precision:** Precision (4) is defined as the ratio between the true positives and the total predicted number of samples that are indicated as positive.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Precision is particularly important in scenarios where false positive predictions carry significant consequences. High precision indicates that when the model predicts a positive class, it is likely to be correct, though it may miss some true positive instances.

- **Recall:** Recall (5) is the ratio of true positives to the total actual number of samples reported as positive. Recall captures the model's ability to identify positive instances and avoid false negative errors. High recall indicates comprehensive detection of positive cases, though it may come at the cost of increased false positives.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

- **F1-score:** The F1-score (6) provides a harmonic mean of precision and recall, offering a balanced measure that considers both false positives and false negatives. The harmonic mean ensures that the F1-score is low when either precision or recall is poor, making it particularly valuable for evaluating model performance when both false positives and false negatives are equally important. The F1-score ranges from 0 to 1, with 1 representing perfect precision and recall. It is computed as:

$$F1\ score = \frac{2\ \times Precision\ \times Recall}{Precision + Recall} \tag{6}$$

The precision, recall, and F1-score were computed using macro-averaging, which calculates metrics independently for each class and then takes the unweighted mean of these values. This approach ensures equal treatment of all classes regardless of their frequency in the dataset. The macro-averaging approach provides insights into model performance across all classes and prevents bias toward majority classes that might occur with micro-averaging. All metrics were computed using the confusion matrix generated from test set predictions, ensuring consistent and standardized evaluation across all algorithms. Comprehensive evaluation using multiple metrics allows for a nuanced interpretation of model performance, as different metrics may favor different aspects of classification behavior. This multi-metric approach enables the identification of models that excel in specific performance dimensions and supports informed decision-making based on the relative importance of different types of classification errors in the given application context.

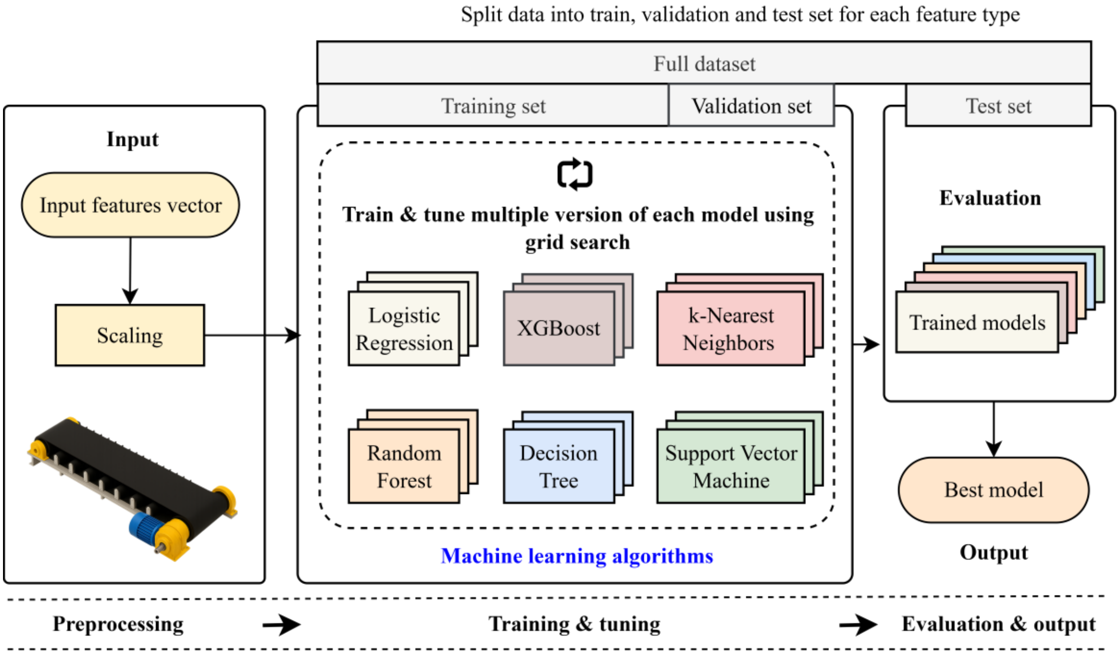


Figure 3 Model development framework for conveyor fault classification

Figure 3 presents a comprehensive overview of the model development framework implemented in this study for conveyor fault classification. The pipeline begins with the input features, followed by preprocessing steps that include feature scaling and partitioning the data into training, validation, and test sets. Six different machine learning classifiers were then trained. Each trained model was evaluated on an unseen test set to determine the most effective solution for fault classification. The best-performing model demonstrates strong generalization capability and is well-suited for deployment in a production environment to enable accurate and reliable fault detection on future data. This structured pipeline in the model development framework ensures a rigorous and reproducible approach to model development, tuning, and evaluation.

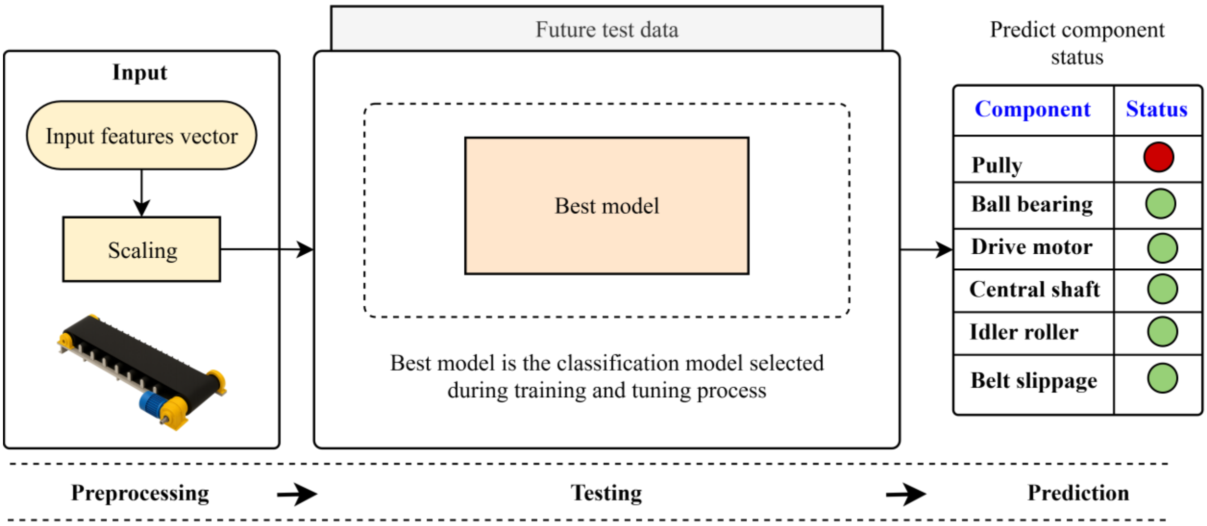


Figure 4 Best model deployment in a production environment for prediction on future data

Figure 4 illustrates how the best classification model can be deployed within a production environment following the training and tuning phases. The first phase (preprocessing) is the same as the training tuning process; the resulting feature vector is then used as an input to the classification model. The model then evaluates the input to predict the status of various components, including the ball bearing, drive motor, central shaft, and belt slippage. The model outputs a status for each component, with a red indicator representing a fault or failure (e.g., faulty ball bearing) and green indicators signifying normal operation. This deployment process enables real-time predictive maintenance and monitoring within an industrial or production setting. The results of the model evaluation, including detailed performance comparisons across all algorithms and metrics, are presented in the following section.

4 Results and discussion

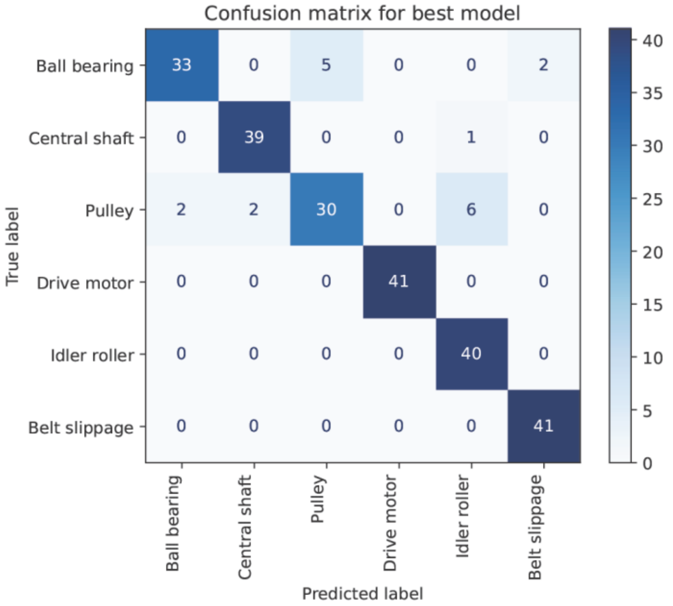
This section presents the outcomes of the machine learning-based conveyor fault classification system developed in this study. The objective was to accurately classify six distinct types of mechanical faults commonly encountered in conveyor systems: ball bearing failure, central shaft issues, pulley malfunction, drive motor faults, idler roller defects, and belt slippage. Various supervised machine learning algorithms were trained and evaluated, and their performance was compared using several standard metrics, including precision, recall, F1-score, MCC, and overall accuracy. These metrics collectively offer a comprehensive understanding of how well each model performs in both detecting and correctly identifying the different fault categories. Table 4 summarizes the classification performance of six different algorithms: Logistic Regression, k-Nearest Neighbors, Random Forest, Decision Tree, Support Vector Machines, and XGBoost. The best performance scores are present in bold font, and the second-best scores are underlined in Table 4. Among these, the Random Forest classifier exhibited the highest overall performance, achieving an accuracy of 92.56%, the highest among all models tested. In addition to accuracy, it also achieved the highest scores across other evaluation metrics, with a precision and recall of 0.93, an F1-score of 0.92, and an MCC of 0.91. These results indicate not only the model's ability to correctly classify the majority of fault instances but also its robustness across all classes, exhibiting a high MCC score. On the other hand, the Logistic Regression model achieved the lowest performance among all classifiers, with an accuracy of 73.55%, a precision and recall of 0.73, and an F1-score of 0.73. These results highlight the limitations of linear models in capturing the complex, nonlinear relationships inherent in conveyor fault patterns. In comparison, XGBoost and Decision Tree classifiers both attained the second-highest accuracy scores of 91.32% and exhibited balanced performance across evaluation metrics. Each achieved a precision, recall, F1-score of 0.91, and MCC of 0.90. Other classifiers, such as k-Nearest Neighbors and Support Vector Machines (SVM), performed reasonably well but were comparatively less effective. KNN achieved an accuracy of 85.12%, with a precision and recall of 0.85 and an MCC of 0.82, while SVM recorded a slightly higher accuracy of 85.54%, with corresponding precision and recall of 0.86 and 0.85, respectively, and an MCC of 0.83. The random forest classifier outperformed all other models across every metric. These results affirm the random forest's capability to effectively model the complex fault patterns across multiple conveyor components and justify its selection as the optimal model for deployment and fault-specific analysis.

Table 4 Various performance scores of different models

Classifier	**Precision**	**Recall**	**F1**	**MCC**	**Accuracy (%)**
Logistic Regression	0.73	0.73	0.73	0.68	73.55
k-Nearest Neighbors	0.85	0.85	0.85	0.82	85.12
Random Forest	**0.93**	**0.93**	**0.92**	**0.91**	**92.56**
Decision Tree	0.91	0.91	0.91	0.90	91.32
Support Vector Machines	0.86	0.85	0.85	0.83	85.54
XGBoost	0.91	0.91	0.91	0.90	91.32

To gain a deeper understanding of how the Random Forest model performs across individual fault categories, the confusion matrix is shown in Figure 4. This confusion matrix presents the classification performance of the Random Forest model across six fault classes. The matrix demonstrates generally strong predictive accuracy, with several notable patterns worthy of detailed examination. The overall diagonal dominance of the confusion matrix indicates that the random forest model successfully learned distinguishing characteristics for each component type. The model exhibits strong performance for certain component types, achieving perfect or near-perfect classification accuracy. Drive motor components show perfect classification with all 41 instances correctly identified (100% accuracy), while idler roller components demonstrate similarly robust performance with 40 out of 40 instances correctly classified. Belt slippage classification also proves highly effective, with all 41 cases accurately predicted, indicating the model's strong capacity to distinguish these particular failure modes.

Ball bearing classification presents a more complex pattern, correctly identifying 33 out of 40 instances (82.5% accuracy). The misclassifications are distributed across pulley (5 instances) and belt slippage (2 instances), suggesting some similarities between these component failure signatures that challenge the model's discriminative capability. Central shaft classification demonstrates strong overall performance with 39 out of 40 instances correctly identified (97.5% accuracy), with only a single misclassification as the idler roller category. This high accuracy indicates well-defined characteristic features for central shaft failures that the random forest algorithm can effectively capture. Pulley classification reveals the most challenging classification task for the model, achieving 30 correct classifications out of 40 total instances (75% accuracy). The error distribution shows misclassifications spread across ball bearing (2 instances), central shaft (2 instances), and, notably, idler roller (6 instances). This pattern suggests potential overlap in the feature space between pulley failures and other mechanical components, particularly idler rollers, which may share similar feature signatures. These results demonstrate the best random forest's robust performance in conveyor fault diagnosis while highlighting specific areas where feature engineering or additional data collection might improve classification accuracy for the more challenging component pairs.



	Ball bearing	Central shaft	Pulley	Drive motor	Idler roller	Belt slippage
Ball bearing	33	0	5	0	0	2
Central shaft	0	39	0	0	1	0
Pulley	2	2	30	0	6	0
Drive motor	0	0	0	41	0	0
Idler roller	0	0	0	0	40	0
Belt slippage	0	0	0	0	0	41

Figure 4 Confusion matrix best model (Random Forest)

Further class-level performance analysis is shown in Figure 5, which compares the precision, recall, and F1-score for each of the fault categories. This performance analysis provides a comprehensive evaluation of the random forest model's classification capabilities across fault classes, revealing distinct performance patterns that align with the confusion matrix findings. The model demonstrates exceptional performance for drive motor and belt slippage classifications, both achieving perfect scores of 1.00 across precision, recall, and F1-score metrics. This indicates complete success in identifying these failure modes without any false positives or missed detections, suggesting these components possess highly distinctive failure signatures that the Random Forest algorithm can readily distinguish. Central shaft classification exhibits robust performance with a precision of 0.95, a recall of 0.98, and an F1-score of 0.96. The slightly higher recall compared to precision indicates the model successfully captures most central shaft failures while maintaining strong specificity, consistent with the minimal misclassifications observed in the confusion matrix. Ball bearing and idler roller classifications present contrasting performance profiles. Ball bearing achieves high precision (0.94) but lower recall (0.83), resulting in an F1-score of 0.88, indicating the model rarely misclassifies other components as ball bearings but occasionally fails to identify actual ball bearing failures. Conversely, the idler roller demonstrates perfect recall (1.00) with lower precision (0.87) and an F1-score of 0.93, suggesting the model captures all idler roller failures but sometimes misclassifies other components as idler rollers. Pulley classification reveals the most challenging discrimination task with a precision of 0.88, a recall of 0.75, and an F1-score of 0.81. The lower recall indicates difficulty in identifying all pulley failures, with some being confused with other components, particularly idler rollers, as shown in the confusion matrix. This suggests potential overlap in failure characteristics between these rotational components.

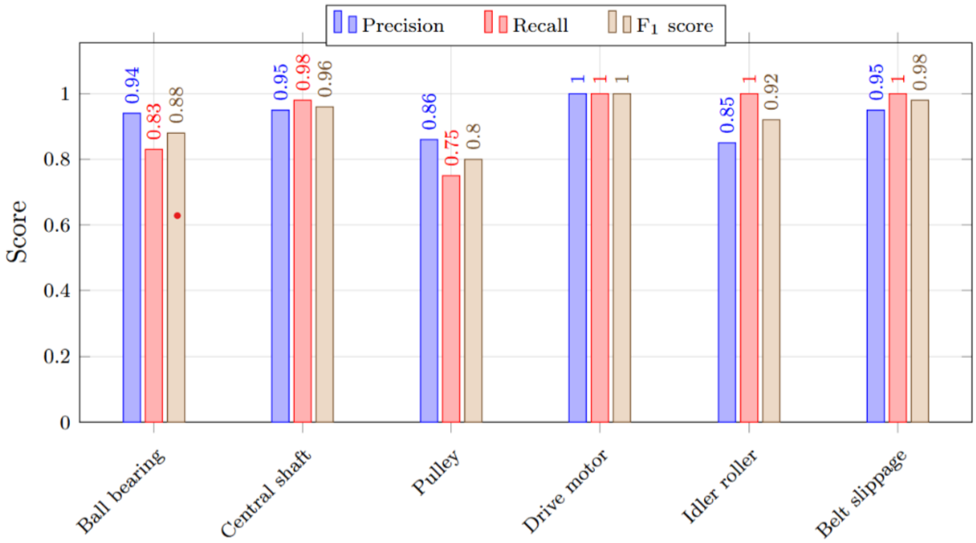


Figure 5 Comparison of class-wise precision, recall, and F1 score of the best model

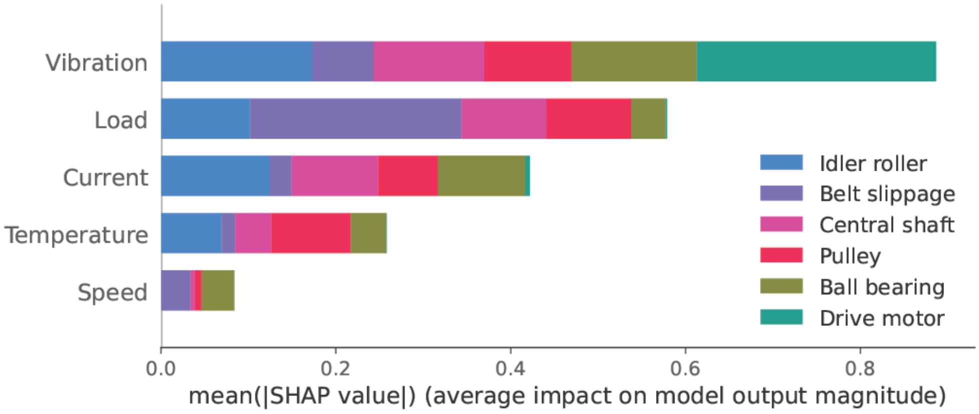


Figure 6 SHAP plot of mean absolute SHAP values across five features for the best model (Random Forest)

Figure 6 shows an SHAP plot explaining the feature contribution towards each fault class. SHAP (SHapley Additive exPlanations) is a model-agnostic interpretability method based on cooperative game theory, where each feature is assigned a Shapley value representing its contribution to a specific prediction. In essence, SHAP quantifies how much each feature increases or decreases the model's output by considering all possible combinations of features. By visualizing these contributions, SHAP helps diagnose model behavior, uncover hidden data relationships, and build trust in machine learning systems by making complex models more transparent. More specifically, the plot shown in Figure 6 presents a detailed visualization of the mean absolute SHAP values across five features for six fault classes in the conveyor system, illustrating the average impact of each feature on the model's predictions for each class. Each horizontal bar corresponds to one of the five features—Vibration, Load, Current, Temperature, and Speed, while the colors within each bar represent the relative contribution of that feature to the prediction of each class. The length of each colored segment within a bar indicates the average contribution (mean absolute SHAP value) of that feature to a specific fault class. The SHAP analysis shows that Vibration and Load are the most influential features, with vibration being especially critical for detecting Drive Motor, Pulley, and Ball Bearing faults, and load being most relevant for Belt Slippage and Central Shaft issues. The current has moderate importance, mainly for Central Shaft, Ball Bearing, and Pulley faults, while temperature is lower in impact but indicative of heat-related problems, particularly in Pulley, Ball Bearing, and Central Shaft. Speed has the least influence, offering only slight contributions to Belt Slippage, Ball Bearing, and Pulley. Overall, the results highlight vibration and load as key diagnostic signals, with current and temperature playing supportive roles and speed having limited predictive value. Overall, the model achieves strong performance, successfully addressing the identified literature gap by demonstrating effective conveyor fault classification capabilities across six distinct fault categories using a unified model. These results validate the feasibility of comprehensive multi-fault classification systems and contribute to the development of more robust condition monitoring solutions for industrial conveyors that can generalize across multiple fault types simultaneously.

5 Conclusion and future work

This study introduces a comprehensive framework for multi-class fault classification in conveyor systems, addressing the limitations of traditional binary approaches by enabling precise fault identification. The framework was validated using real-world operational data collected under variable loading conditions, ensuring practical relevance for logistics operations. Among the evaluated algorithms, Random Forest demonstrated superior performance, achieving an accuracy of 92.56%, a precision of 0.93, a recall of 0.93, an F1-score of 0.92, and an MCC of 0.91. Beyond predictive accuracy, the model's interpretability was enhanced through SHAP-based feature analysis, which provided insights into the most influential diagnostic signals for each fault class. These capabilities support informed maintenance decisions, operator training, and seamless integration with Industry 4.0 initiatives, contributing to proactive and cost-effective asset management for logistic applications.

Despite these strengths, certain fault types, such as pulley faults, remain challenging due to overlapping feature spaces with mechanically similar components. This limitation underscores the need for further refinement and motivates future research. Promising directions can be grouped under three broader themes: (i) Integrating complementary modalities such as acoustic signals, vibration patterns, and thermal imaging to enrich the feature space and improve classification accuracy for complex fault scenarios; (ii) Investigating deep learning architectures, including CNNs for spatial feature extraction and LSTMs for temporal dependency modeling, to capture nuanced fault progression patterns directly from raw sensor data; (iii) Developing edge-computing solutions for on-site inference, coupled with extensions for fault severity assessment and remaining useful life prediction, to deliver actionable insights and enable predictive maintenance strategies. By addressing these directions, the proposed framework can evolve into a robust, scalable solution for intelligent condition monitoring and decision support in modern logistics environments.

References

[1] ALHARBI, F., LUO, S.: A review of fault detecting devices for belt conveyor idlers, *Journal of Mechanical Engineering Science and Technology*, Vol. 8, No. 1, pp. 39-53, 2024. https://doi.org/10.17977/um016v8i12024p039

[2] WIJAYA, H., RAJEEV, P., GAD, E., VIVEKANANTHAM, R.: Automatic fault detection system for mining conveyor using distributed acoustic sensor, *Measurement*, Vol. 187, 110330, pp. 1-18, 2022. https://doi.org/10.1016/j.measurement.2021.110330

[3] ALHARBI, F., LUO, S., ZHANG, H., SHAUKAT, K., YANG, G., WHEELER, C.A., CHEN, Z.: A Brief Review of Acoustic and Vibration Signal-Based Fault Detection for Belt Conveyor Idlers Using Machine Learning Models, *Sensors*, Vol. 23, No. 4, 1902, pp. 1-33, 2023. https://doi.org/10.3390/s23041902

[4] MILOVANCEVIC, M.: Diagnosing Belt Conveyor Idler Faults with STFT and CNN, *International Journal of Acoustics and Vibration*, Vol. 29, No. 4, pp. 494-501, 2024. https://doi.org/10.20855/ijav.2024.29.42091

[5] SKOCZYLAS, A., STEFANIAK, P., ANUFRIIEV, S., JACHNIK, B.: Belt conveyors rollers diagnostics based on acoustic signal collected using autonomous legged inspection robot, *Applied Sciences*, Vol. 11, No. 5, 2299, pp. 1-13, 2021. https://doi.org/10.3390/app11052299

[6] WODECKI, J., SHIRI, H., SIAMI, M., ZIMROZ, R.: *Acoustic-based diagnostics of belt conveyor idlers in real life mining conditions by mobile inspection robot*, In: Conference on Noise and Vibration Engineering, Leuven, Belgium, pp. 535-543, 2022.

[7] LIU, X., PEI, D., LODEWIJKS, G., ZHAO, Z., MEI, J.: Acoustic signal based fault detection on belt conveyor idlers using machine learning, *Advanced Powder Technology*, Vol. 31, No. 7, pp. 2689-2698, 2020. https://doi.org/10.1016/j.apt.2020.04.034

[8] ZHANG, M., ZHANG, Y., ZHOU, M., JIANG, K., SHI, H., YU, Y., HAO, N.: Application of Lightweight Convolutional Neural Network for Damage Detection of Conveyor Belt, *Applied sciences*, Vol. 11, No. 16, 7282, pp. 1-15, 2021. https://doi.org/10.3390/app11167282

[9] LIU, M., ZHU, Q., YIN, Y., FAN, Y., SU, Z., ZHANG, S.: Damage detection method of mining conveyor belt based on deep learning, *IEEE Sensors Journal*, Vol. 22, No. 11, pp. 10870-10879, 2022. https://doi.org/10.1109/JSEN.2022.3170971

[10] DWIVEDI, U.K., KUMAR, A., SEKIMOTO, Y.: *Realtime damage detection in long conveyor belts using deep learning approach*, In: 40th Proceedings of the International Symposium on Automation and Robotics in Construction, Chennai, India, pp. 10-115, 2023. https://doi.org/10.22260/ISARC2023/0017

[11] GUNCKEL, P.V., LOBOS, G., RODRIGUEZ, F.K., BUSTOS, R.M., GODOY, D.: Methodology proposal for the development of failure prediction models applied to conveyor belts of mining material using machine learning, *Reliability Engineering & System Safety*, Vol. 256, 110709, pp. 1-15, 2025. https://doi.org/10.1016/j.ress.2024.110709

[12] BORTNOWSKI, P., NOWAK-SZPAK, A., OZDOBA, M., KROL, R.: The acoustic camera as a tool to identify belt conveyor noises, *Journal of Sustainable Mining*, Vol. 19, No. 4, pp. 286-294, 2020. https://doi.org/10.46873/2300-3960.1036

[13] FEDORKO, G., LIPTAI, P., MOLNAR, V.: Proposal of the methodology for noise sources identification and analysis of continuous transport systems using an acoustic camera, *Engineering Failure Analysis*, Vol. 83, pp. 30-46, 2018. https://doi.org/10.1016/j.engfailanal.2017.09.011

[14] YANG, W., ZHANG, X., MA, H.: *An inspection robot using infrared thermography for belt conveyor*, In: IEEE 13th International Conference on Ubiquitous Robots and Ambient Intelligence, Xi'an, China, pp. 400-404, 2016. https://doi.org/10.1109/URAI.2016.7734069

[15] SIAMI, M., BARSZCZ, T., WODECKI, J., ZIMROZ, R.: Automated Identification of Overheated Belt Conveyor Idlers in Thermal Images with Complex Backgrounds Using Binary Classification with CNN, *Sensors*, Vol. 22, No. 24, pp. 10004, 2022. https://doi.org/10.3390/s222410004

[16] ZHANG, M., CAO, Y., JIANG, K., LI, M., LIU, L., YU, Y., ZHOU, M., ZHANG, Y.: Proactive measures to prevent conveyor belt failures: Deep learning-based faster foreign object detection, *Engineering failure analysis*, Vol. 141, 106653, pp. 1-14, 2022. https://doi.org/10.1016/j.engfailanal.2022.106653

[17] XIUYU, Y., LIYONG, T., FENG, C.: Thermal infrared imaging for conveyor roller fault detection in coal mines, *Public Library of Science*, Vol. 19, No. 7, e0307591, pp. 1-17, 2024. https://doi.org/10.1371/journal.pone.0307591

[18] SIAMI, M., BARSZCZ, T., WODECKI, J., ZIMROZ, R.: Semantic segmentation of thermal defects in belt conveyor idlers using thermal image augmentation and u-net-based convolutional neural networks, *Scientific Reports*, Vol. 14, No. 1, 5748, pp. 1-15, 2024. https://doi.org/10.1038/s41598-024-55864-2

[19] LI, X., LI, Y., ZHANG, Y., LIU, F., FANG, Y.: Fault diagnosis of belt conveyor based on support vector machine and grey wolf optimization, *Mathematical Problems in Engineering*, Vol. 2020, No. 1, 1367078, pp. 1-10, 2020. https://doi.org/10.1155/2020/1367078

[20] SOARES, J.L., COSTA, T.B., MOURA, L.S., SOUSA, W.S., MESQUITA, A.L., MESQUITA, A.L., DE FIGUEIREDO, J.M., BRAGA, D.S.: Fault diagnosis of belt conveyor idlers based on gradient boosting decision tree, *The International Journal of Advanced Manufacturing Technology*, Vol. 132, No. 7, pp. 3479-3488, 2024. https://doi.org/10.1007/s00170-024-13549-0

[21] LIU, J., FU, S., LIU, F., CHENG, X.: Intelligent fault diagnosis of belt conveyor rollers using a polar KNN algorithm with audio features, *Engineering Failure Analysis*, Vol. 168, 109101, pp. 1-13, 2025. https://doi.org/10.1016/j.engfailanal.2024.109101

Review process
Single-blind peer review process.