ASEAN
Journal of Scientific and Technological Reports
Online ISSN:2773-8752

Research article

Designing and Developing Quality Control of Processes Using the Failure Mode and Effects Analysis Method and Machine Learning

Sirirat Pungchompoo[1*], Nikorn Sirivongpaisal[2], Rakkrit Duansoithong[3], and Aree Teeraparbseree[4]

[1] Faculty of Engineering, Prince of Songkla University, Songkhla, 90112, Thailand
[2] Faculty of Engineering, Prince of Songkla University, Songkhla, 90112, Thailand
[3] Faculty of Engineering, Prince of Songkla University, Songkhla, 90112, Thailand
[4] Faculty of Engineering, Prince of Songkla University, Songkhla, 90112, Thailand
* Correspondence: sirirat.pu@psu.ac.th

Citation:
Pungchompoo, S.; Sirivongpaisal, N.; Duansoithong, R.; Teeraparbseree, A. Designing and developing quality control of processes using the failure mode and effects analysis method and machine learning. *ASEAN J. Sci. Tech. Report.* **2025**, *28*(6), e259345. https://doi.org/10.55164/ajstr.v28i6.259345.

Article history:
Received: May 15, 2025
Revised: September 15, 2025
Accepted: September 27, 2025
Available online: October 20, 2025

Abstract: Medium-density fiberboard (MDF) production involves multiple intricate stages. Uneven thickness formation, large volumes of dynamic data from automated systems, and rapidly evolving technologies create substantial challenges for maintaining consistent quality control. Conventional approaches rely heavily on expert judgment and lack predictive capability, leaving a critical gap in timely and accurate risk assessment. This study addresses these challenges by integrating Failure Mode and Effects Analysis (FMEA) with machine learning techniques to evaluate and predict risks throughout the MDF production process. Real production data from an industrial facility were used to ensure practical relevance. Domain experts first assessed the Severity (S), Occurrence (O), and Detection (D) parameters using the PFMEA method. Predictive models—including K-Nearest Neighbors, Support Vector Machine, Neural Network, and an Ensemble Method—were then developed to estimate risk scores. The findings show that the Neural Network and Ensemble Method achieved the highest overall accuracy. This integrated approach reduces subjective bias, enhances predictive precision, and supports informed decision-making for quality control and risk mitigation in industrial MDF production.

Keywords: Failure mode and effects analysis (FMEA); machine learning; data driven quality control

1. Introduction

The medium-density fiberboard (MDF) industry plays a significant role in the production of furniture and construction materials. In Thailand, the abundance of rubberwood makes it an ideal raw material for MDF manufacturing. Despite its potential, industrial MDF production continues to face challenges related to quality control, real-time data processing from automated systems, and adaptation to emerging technologies within the context of Industry 4.0. A case study identified inconsistency in board thickness as a major issue, mainly caused by fluctuations in temperature and pressure during the continuous hot press (CHP) process. This inconsistency reduces product quality, increases production costs, and affects customer confidence. Moreover, the large volume of complex, time-dependent data generated by modern manufacturing equipment makes it difficult to conduct timely analysis without an effective risk assessment system.

Failure Mode and Effects Analysis (FMEA) is a widely used method for identifying and ranking potential risks in manufacturing processes. It uses a Risk Priority Number (RPN), calculated based on severity (S), occurrence (O), and detection (D), to help manage quality problems. However, traditional FMEA has some limitations. It is typically static, heavily dependent on expert opinions, and cannot adapt to real-time changes. As a result, many recent studies have combined FMEA with other techniques, such as Statistical Process Control (SPC), fuzzy logic, and artificial intelligence (AI), to improve its effectiveness.

For example, one study introduced a dynamic FMEA model using data from CNC machine operations and maintenance to reduce reliance on expert judgment and increase risk evaluation accuracy [1]. In another case, combining SPC with FMEA helped detect severe defects in hollow steel production, where poor machine setup led to an RPN of 360 [2]. In the chemical industry, this integration reduced process variation by 63% [4], while in furniture manufacturing, the use of Fuzzy-FMEA with SPC revealed that worker inexperience was the primary cause of product defects [3]. To overcome the limitations of traditional FMEA, machine learning methods are being increasingly utilized. One study applied a Fuzzy Adaptive Resonance Theory (Fuzzy ART) model to classify failure modes without fixed thresholds [5]. Another developed a fuzzy-based FMEA that used expert input and linguistic terms. This model was applied to a Load-Haul-Dump (LHD) machine in underground mining and showed that the electrical system had the highest fuzzy RPN of 117 [6]. AI-based methods have also widened the use of FMEA. For instance, chatbot-driven Social FMEA was used to assess socially responsible product designs [7]. In sustainable construction, a hybrid framework combining Artificial Neural Networks (ANN), fuzzy logic, and the Internet of Things (IoT) reached 92.7% prediction accuracy and adapted well to noisy and changing data [8]. In addition, optimization methods also help improve risk assessment. A two-stage optimization model for household energy systems lowered user costs by 30% and increased energy efficiency by 4.8% [9]. Furthermore, Random Forest has outperformed traditional logistic regression in several fields, including education [10] and political studies [11]. Logistic regression has also helped rank risks in solar power plants [12]. Finally, linking IoT with FMEA is a promising trend. Real-time sensor data, such as temperature, humidity, thickness, and pressure, can update RPN scores and support predictive models. This helps build proactive, data-driven quality control systems that meet the complex needs of today's production environments.

This highlights a clear research gap. Although Fuzzy FMEA and SPC-FMEA have been applied in risk assessment, both approaches show notable limitations in the context of MDF manufacturing, where real production data are complex, non-linear, and imbalanced. Fuzzy FMEA helps reduce ambiguity in expert scoring but relies on predefined membership functions and rule sets, which makes it challenging to adapt to changing data conditions. SPC-FMEA effectively tracks statistical fluctuations but lacks the flexibility to predict S, O, and D in advance. In contrast, integrating PFMEA with machine learning enables the model to learn directly from actual production data, handle class imbalance and non-linear relationships more effectively, and predict S, O, and D with greater accuracy. This integration enables systematic and timely RPN calculation and risk prioritization, particularly well-suited to the challenges of MDF production.

Therefore, this study aims to develop a data-driven quality control model for MDF manufacturing by integrating FMEA, IoT sensor data, and machine learning algorithms. The conceptual framework consists of two main components. The first focuses on identifying risk factors using FMEA and developing predictive models with methods such as Random Forest, XGBoost, and Neural Networks. The second introduces an Inverted Process Model that utilizes predictive outputs to recommend optimal process parameters, reduce the RPN, and enhance product quality. As illustrated in Figure 1, the proposed framework comprises three main stages: (1) Data Collection from automated systems, (2) Data Access and Preparation, and (3) Smart Data and Modeling through machine learning. By combining risk analysis with intelligent prediction, this approach facilitates the development of an intelligent quality control system that aligns with Industry 4.0.

2. Materials and Methods

This study aims to develop an integrated approach for analyzing and forecasting risks in the production process of medium-density fiberboard (MDF). The method combines expert-based Failure Mode and Effects Analysis (FMEA) with quantitative techniques, including statistical analysis and machine learning algorithms. The objective is to improve the accuracy of assessing severity (S), occurrence (O), and detection (D), which are then used to calculate the Risk Priority Number (RPN) systematically. The outcome is to implement the model in a web application designed for practical use in industrial settings. The research methodology consists of two main stages.

Stage 1: Risk Evaluation Using PFMEA Based on Expert Judgments

The initial risk assessment used the Process FMEA (PFMEA) method, drawing on evaluations from three to five experts at an MDF manufacturing facility. These experts—engineers, production supervisors, and quality control staff—assigned scores independently for S, O, and D to potential failure modes within the production process. Scoring followed the standard FMEA 1–10 scale: Severity (S) ranged from 1 (negligible effect on product quality) to 10 (catastrophic failure affecting safety or compliance); Occurrence (O) from 1 (failure unlikely to occur) to 10 (failure almost inevitable to occur); and Detection (D) from 1 (failure almost inevitable to be detected) to 10 (failure improbable to be detected). The scores were averaged across experts, and RPN values were calculated (RPN = S × O × D). These data were then used to prioritize risk areas and provided the basis for developing predictive machine learning models in the next phase.

Stage 2: Development and Evaluation of Machine Learning Models

To address the non-linear and imbalanced nature of real production data, machine learning techniques were applied to predict S, O, and D values. The dataset was divided into 80% for training and 20% for testing. Within the training set, stratified 5-fold cross-validation (k = 5) was used to tune and evaluate models while preserving class proportions. No repeated cross-validation was applied. All variables were standardized and normalized prior to training to minimize the effects of different units and scales.

Because the data were complex and imbalanced across classes, oversampling techniques (RandomOverSampler/SMOTE) were applied only to the training set, with the test set kept untouched to preserve the original distribution. The Ensemble Method, combined with Bagging and Decision Trees, was also selected to reduce bias and variance, as well as to handle class imbalance effectively.

Severity (S), Occurrence (O), and Detection (D) were classified into 10 classes (Class 1–Class 10) following the standard FMEA scoring scale. The actual production data showed an imbalanced distribution—for example, S had the highest frequency in Class 1 (103 records) and fewer in higher classes; O was concentrated in Classes 1–3 (100–103 records); and D peaked in Class 6 (215 records) with some classes having no records. This imbalance was explicitly taken into account in model selection and data preprocessing.

Four algorithms were selected for model development, with the following key settings: K-Nearest Neighbors (KNN), k = 10 neighbors; Euclidean distance metric; squared inverse distance weight; and standardized data. Support Vector Machine (SVM): cubic kernel; kernel scale automatic; box constraint level 1; multiclass coding one-vs-one; standardized data. Neural Network (NN): one fully connected hidden layer; 100 neurons; ReLU activation; iteration limit 1000; regularisation strength (lambda) 0; standardized data. Ensemble Method (Bagging + Decision Trees): 30 learners; decision tree base learners; maximum number of splits 479; all predictors sampled.

Model development included data preprocessing, training, and tuning with stratified 5-fold cross-validation on the training set, followed by final performance evaluation on the 20% hold-out test set. Performance was assessed using classification metrics—Accuracy, Balanced Accuracy, Precision, Recall, and Macro-F1 Score—as well as regression-style indicators (R^2, RMSE, MAE) for comparison. A feature importance analysis was also conducted to identify key factors influencing risk. The predicted S, O, and D values were then used to compute new RPN scores, which were applied to guide process improvements and risk mitigation efforts.

The PFMEA focused specifically on four upstream processes that precede the Refiner stage, where initial observations indicated premature blade damage. This issue appeared to be caused by upstream conditions, including raw material contamination and technical inconsistencies. Therefore, four critical production processes were selected for detailed analysis.

- Debarking involves removing bark and foreign materials from logs. The process begins with log alignment using a vibration conveyor, followed by the removal of stones and metals via a roller stone trap. Logs are then transferred through handling systems to the debarking drum, which removes 60–80% of the bark.
- Chipping converts debarked logs into wood chips within a size range of 4–40 millimeters. The material is conveyed through a chain conveyor, bark separator, and belt conveyor to the chipper. Blade performance is continuously monitored through current (ampere) measurements to prevent damage.
- Chip washing reduces impurities and moisture in the wood chips. Chips are moved using a twin-screw system through a roller screen to classify size, washed with water, and dewatered using a screw press and plug screw. The moisture content is finally adjusted in a surge bin to meet the process requirements.
- Digesting softens the wood chips using high-pressure steam. The digester regulates both chip size and steam pressure within optimal ranges, ensuring the material is conditioned correctly before entering the refiner.

3. Results and Discussion

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, and the conclusions that can be drawn from the experiment. Authors should discuss the results and how they can be interpreted in light of previous studies and the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

3.1 PFMEA Assessment

The risk analysis of the MDF production process, conducted prior to the Refiner stage using the PFMEA method, revealed that the primary causes of premature Refiner blade wear are related to the presence of contaminants, unsuitable moisture levels, and a lack of effective defect detection, which still relies primarily on visual inspection. The PFMEA results are illustrated in Figure 1. During the debarking process, issues were observed with the vibration conveyor and roller conveyor trap, where wood stacking and the presence of foreign materials, such as stones and soil, frequently interrupted the material flow. This resulted in a high RPN score of 128. In the chipping process, particularly involving the chain conveyor, accumulated bark often caused the chipper to stop, leading to the highest recorded RPN of 192. This was attributed to the severity of the issue and the absence of an automated detection system. In the chip washing stage, residual contaminants and excessive moisture levels were found. These problems were linked to ineffective washing and dewatering operations, resulting in RPN values ranging from 120 to 126. These factors were shown to contribute to blade wear in the refiner. Overall, the key risks identified were related to poor control of contamination, suboptimal moisture management, and inaccurate detection methods. These findings suggest the need for process improvement through the implementation of automated sensor systems and enhanced filtering and conveying mechanisms to minimize equipment damage and stabilize the production process.

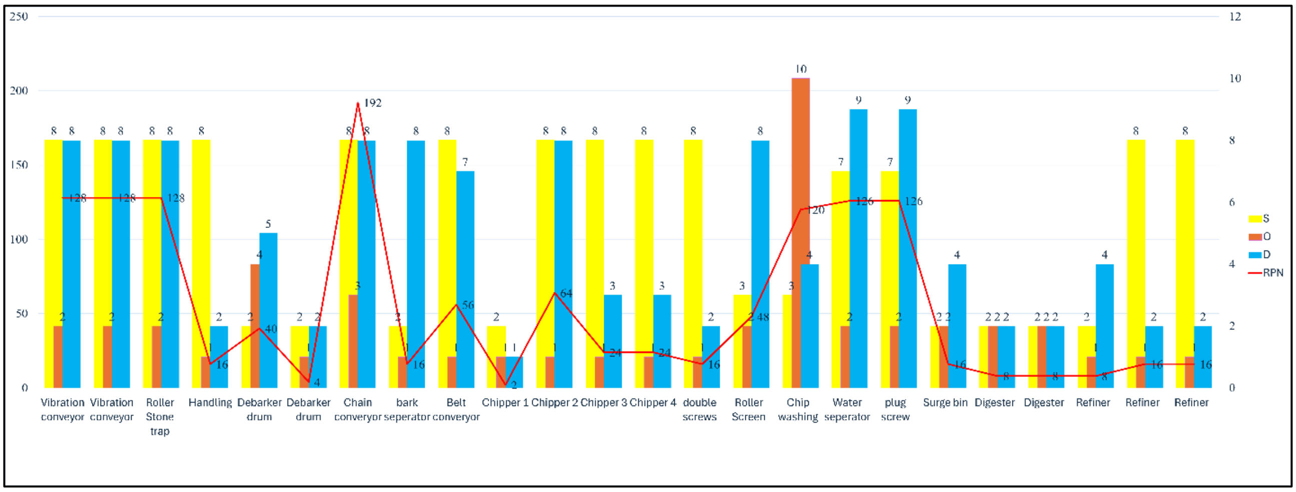

Figure 1. PFMEA Assessment Results

3.2.Development of Linear Models and Statistical Analysis

To validate the results of the PFMEA assessment, a statistical analysis was conducted using 354 actual production records collected between January 2023 and mid-May 2024. The study focused on the wear of the Refiner blade (Wear) and the specific energy consumption (SEC), which is directly associated with wear, in relation to several independent variables. These included chip types (Fine Chip, Good Chip, and Oversize Chip), the amount of bark in the process, both during debarking and before entering the roller screen, as well as the moisture content.

As illustrated in the matrix plot in Figure 2, the analysis revealed clear correlations between these variables and blade wear. The statistical results demonstrated that moisture content was significantly correlated with both the wear of the Refiner blade and the SEC values. Notably, the correlation coefficient between SEC and Wear reached a value of 0.908, indicating a strong direct relationship between energy consumption and the degree of blade deterioration during the production process. Moreover, moisture content was also found to be significantly related to chip type—particularly Fine Chip and Good Chip—as well as to the presence of bark contaminants during the debarking stage and prior to the roller screen. These findings reinforce the PFMEA results, which identified contamination and improper moisture levels as primary causes of equipment wear before the refining process. Additionally, the study confirmed that the current reliance on visual inspection for anomaly detection is insufficient for early identification and prevention of damage.

In summary, the results indicate that moisture content, chip type, and the level of contamination in both conveying and washing processes are key factors contributing to Refiner blade wear. The integration of statistical data analysis with PFMEA evaluation confirms that the development of automated detection systems and data-driven predictive models is essential for minimizing damage and systematically improving the efficiency and reliability of MDF production processes.

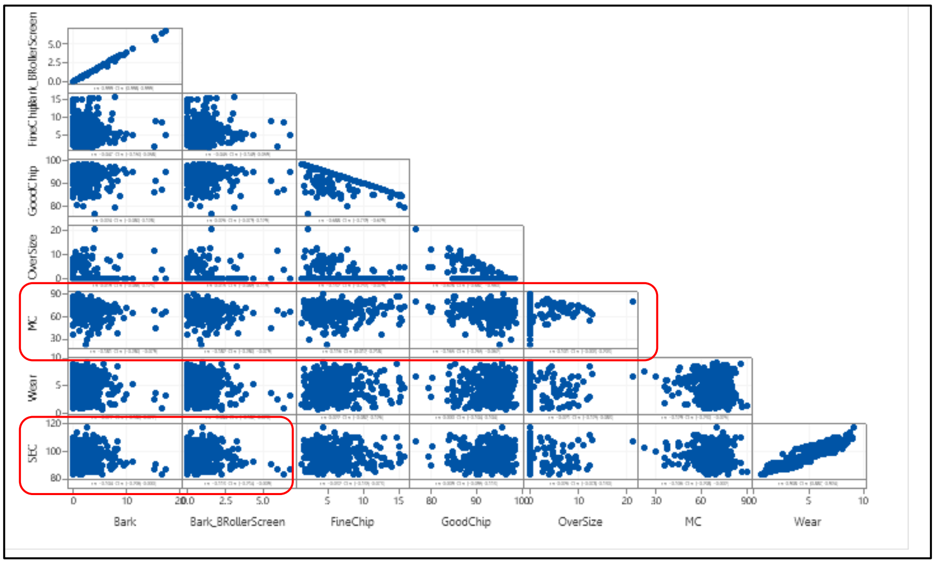

Figure 2. Matrix Plot for Pair Pearson Correlation Analysis

Under the variables analyzed through the PFMEA process, the values of severity (S), occurrence (O), and detection (D) were used as key factors in developing a predictive model for the Risk Priority Number (RPN) through multiple linear regression (MLR). The resulting MLR equation used to forecast the RPN is as follows:

RPN = -66.18 + 28.29 * O + 22.32 * S + 21.18 * D + 1.76 * 2.5<Wear<4 + 1.40 * Over size <5% + 0.28 * 80<SEC<130 + -0.45 * Bark + -0.48 * %MC < 80% + -0.48 * Good chip >80% + -1.76 * Fine chip <15% + -10.70 * Chip Washing_pH < 5 + -10.91 * Chip Washing_Solid < 7

However, the implementation of PFMEA evaluation and the subsequent statistical validation revealed several limitations. While Pearson correlation analysis provided an initial indication of linear relationships among variables, it was insufficient for capturing deeper or non-linear associations. Similarly, the use of

Multiple Linear Regression (MLR) presented constraints in accurately predicting RPN values from real-world production data, which are often complex and asymmetrical in nature. Therefore, it is necessary to develop more flexible machine learning models capable of accurately predicting the values of severity (S), occurrence (O), and detection (D), while minimizing the bias typically associated with expert-based assessments.

3.3 Risk Analysis and Prediction in Manufacturing Processes Using a Data-Driven Approach

The dataset used in this study consists of ten independent variables that reflect the characteristics of wood chips and key stages of the MDF production process, all of which are related to the wear and degradation of equipment in the production line. These variables were employed to develop predictive models and conduct systematic and practical risk analysis, focusing on their relationship with blade wear and overall process efficiency. Detailed descriptions of the variables are presented in Table 1.

Table 1. Key Variables Used in the Study

Variable	Definition
Fine chip <15%	Proportion of fine wood particles, accounting for less than 15% of total chips.
Good chip >80%	Proportion of high-quality wood chips exceeding 80% of the total chip volume.
Over size <5%	Proportion of oversized wood chips, not exceeding 5% of the total.
Bark	Amount of bark separated from the wood chips.
Bark Before Roller Screen	Bark is removed before the screening process using a roller screen.
%MC <80%	Moisture content in wood chips is maintained below 80%.
Chip Washing_pH <5	pH level of the water used in chip washing, where pH is less than 5.
Chip Washing_Solid <7	The concentration of solid particles in chip washing water is maintained at a level below 7%.
2.5 < Wear < 4	Degree of equipment wear ranging between 2.5 and 4.
80 < SEC < 130	Specific Energy Consumption (SEC) within the range of 80 to 130.

The dataset used in this study consists of ten independent variables, each with a considerably different range of values. For instance, some variables have values below 10, while others exceed 100. Without applying proper scaling or standardization prior to analysis, such disparities could adversely affect the performance of predictive models. Moreover, the dataset exhibits a non-linear relationship among variables and a notable degree of class complexity. As illustrated in Figure 3, the scatter plot shows overlapping distributions among Classes 1, 2, 3, and 4, indicating a high level of difficulty in distinguishing between these classes. In contrast, Class 8 is clearly separated from the others, demonstrating a distinct distribution pattern. These findings highlight the challenges in class classification and emphasize the importance of appropriate data preprocessing techniques.

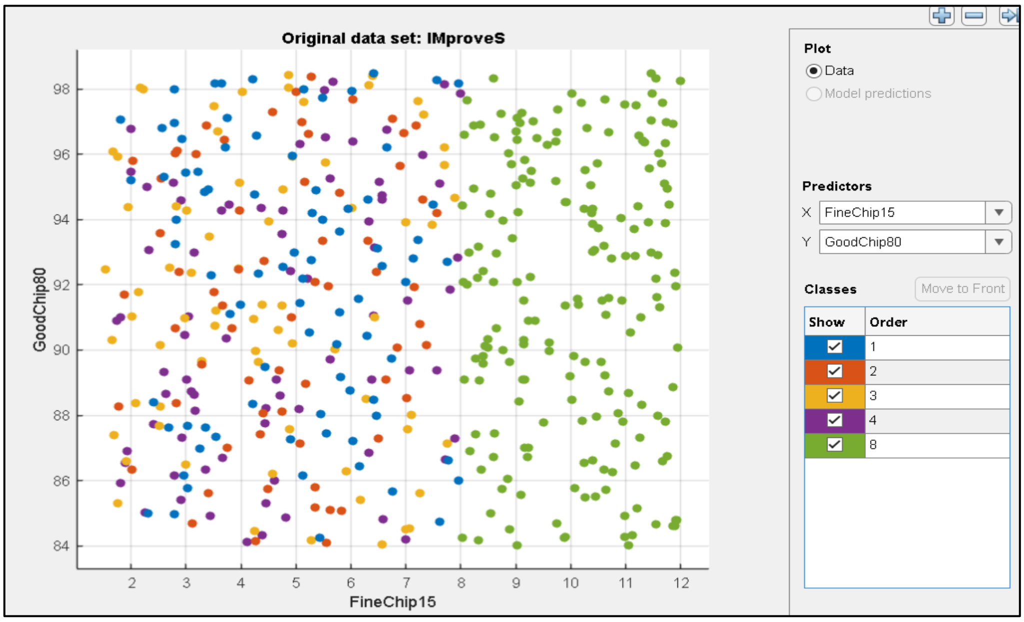

Figure 3. Scatter plot showing the distribution of data across different classes

For this reason, it was necessary to select models capable of handling non-linear relationships, varying data scales, and class complexity. Accordingly, four suitable machine learning models were considered, along with appropriate hyperparameter settings. This study involved the analysis and development of predictive models for risk assessment in MDF manufacturing, based on the principles of Failure Mode and Effects Analysis (FMEA) and using actual industrial data. The dataset presented notable challenges, including non-linearity, class overlap, and variation in the range of independent variables—especially within the S = 1–4 class group, which proved difficult to classify using conventional methods. To address these limitations, four machine learning algorithms were developed: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Neural Network (NN), and an Ensemble Method (Bagging with Decision Trees). Prior to model training, appropriate data preprocessing was performed, including hyperparameter tuning and standardization, to enhance model performance. The effectiveness of each model in predicting the three FMEA indicators—Severity (S), Occurrence (O), and Detection (D)—was evaluated using accuracy scores derived from the test datasets. Because the dataset was imbalanced, the Precision, Recall, and F1-Score reported in the tables are Weighted Averages calculated according to the proportion of samples in each class. The model achieving the highest accuracy for each target variable was identified and compared to determine which model performed best across all three dimensions.

For Severity (S) prediction, K-Nearest Neighbor (KNN) achieved the highest accuracy of 85.8%, followed closely by the Neural Network model at 85.0%. The Ensemble Method achieved an accuracy of 84.2%, while the Support Vector Machine (SVM) yielded the lowest accuracy at 83.3% (Table 2). In addition to accuracy, KNN also yielded substantial weighted precision (86.4%), recall (85.8%), and F1-score (85.5%), indicating consistent classification performance across severity levels, as shown in Table 2.

Table 2. Accuracy of Models in Predicting Severity (S)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
K-Nearest Neighbor	85.8	86.4	85.8	85.5
Neural Network	85.0	87.9	87.5	875
Ensemble Method	84.2	83.7	83.3	83.4
Support Vector Machine	83.3	84.5	83.3	83.5

In terms of Occurrence (O) prediction, the Neural Network model demonstrated the highest accuracy at 81.7%, followed by the Ensemble Method at 79.2%. SVM produced an accuracy of 76.7%, while KNN yielded the lowest accuracy at 75.8% (Table 3). Although the Ensemble Method achieved slightly lower accuracy than the Neural Network, it maintained relatively high weighted precision (81.5%) and recall (81.7%), resulting in a weighted F1-score of 81.2%, as shown in Table 3.

Table 3. Accuracy of Models in Predicting Occurrence (O)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
K-Nearest Neighbor	75.8	77.3	77.5	77.2
Neural Network	81.7	79.8	80.0	79.8
Ensemble Method	79.2	81.5	81.7	81.2
Support Vector Machine	76.7	75.7	75.0	75.2

For Detection (D) prediction, the SVM model achieved the highest accuracy at 85.0%, outperforming the Ensemble Method (80.8%), the Neural Network (80.8%), and KNN (79.2%) (Table 4). While SVM led in overall accuracy, the Ensemble Method showed the most balanced weighted metrics—precision (82.5%), recall (82.5%), and F1-score (82.3%)—reflecting consistent detection capability across classes, as shown in Table 4.

Table 4. Accuracy of Models in Predicting Detection (D)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
K-Nearest Neighbor	79.2	80.0	79.2	79.3
Neural Network	80.8	82.1	81.7	81.7
Ensemble Method	80.8	82.5	82.5	82.3
Support Vector Machine	85.0	83.3	80.8	81.2

3.4 Analysis and Prediction of Risk in MDF Manufacturing Using a Data-Driven Approach

The findings indicated that no single model achieved the highest accuracy across all target variables. However, when considering the overall performance, the Neural Network model consistently produced high accuracy levels and achieved the best results in predicting the occurrence (O) variable. Similarly, the Ensemble Method showed the highest accuracy in predicting detection (D) and performed comparably to the Neural Network in predicting both S and O. In contrast, although the KNN model delivered strong performance in predicting severity (S), it had the lowest accuracy in predicting occurrence and detection, highlighting its limitations in broader applications. Meanwhile, the Support Vector Machine (SVM) demonstrated slightly lower accuracy compared to other models, yet maintained consistent and acceptable results. Therefore, it can be concluded that both Neural Network and Ensemble Method models show strong potential for practical use in predicting all three FMEA indicators and should be further improved through hyperparameter tuning, data augmentation, and class balancing to increase overall prediction accuracy.

3.5 Recommendations for Improving MDF Production to Minimize Risk and RPN

Based on the PFMEA evaluation and data analysis from an actual manufacturing facility, the major risk factors contributing to Refiner blade wear and high RPN values include the proportion of fine chips, moisture content of raw materials, pH control in the chip washing process, and blade wear itself. These findings were supported by statistical analyses and machine learning model results, which identified these variables as significantly influencing severity (S), occurrence (O), and detection (D) scores. Thus, the following recommendations are proposed to improve the production process:

- Control of Fine Chips (FineC_weight) To maintain chip size within standard ranges, it is recommended to optimize cutting speed and feed rate in the Chipper process. Implementing a preventive maintenance plan can also help minimize blade wear, a major cause of uneven cutting and excessive fine chips, thereby reducing the risk in subsequent steps.
- Moisture Control Automatic moisture control systems should be installed in production areas, such as dehumidifiers or temperature regulation systems, to keep moisture levels within acceptable limits. Regular moisture measurements should be carried out using sensors and automated alerts. In addition, raw material storage areas should be improved to ensure better environmental control.
- pH Control in Chip Washing Frequent monitoring of pH levels in the washing process is necessary, and appropriate chemical adjustments should be made when values deviate from acceptable ranges. Effective control of wash water quality can help reduce residual substances that cause equipment damage and compromise chip quality.
- Reduction of Blade Wear (Wear). Scheduled tool maintenance, along with the use of high-durability materials, is essential. Additionally, the use of high-precision cutting equipment can help reduce the occurrence of oversized chips and improve the overall efficiency of the production process.
- Enhanced Detection Systems PFMEA results indicated that visual inspection has significant limitations. The development of automated detection systems using sensors—such as vibration, moisture, and electric current—can enhance real-time problem analysis and reduce the risk of human error.
- Use of Statistical Process Control (SPC) Control charts should be implemented to monitor the stability of the process after improvements. This approach enables real-time monitoring of deviations in the production process, thereby minimizing damage and enhancing long-term product quality.

3.6 Discussion

The risk analysis of the MDF production process using PFMEA, supported by actual operational data, revealed that the primary factors contributing to premature wear of critical equipment—particularly the Refiner blades—were contamination (e.g., stones, soil, and bark), excessive moisture content, and limited fault detection, which still relied mainly on visual inspection. High RPN scores reinforced these findings in several key processes, including the chipping stage with the chain conveyor (RPN = 192), the Debarking stage (RPN = 128), and the Chip Washing stage (RPN = 120–126). Statistical analysis using Pearson correlation confirmed significant positive relationships between moisture content and both blade wear and specific energy consumption (SEC), an indicator of process efficiency. Moisture content also correlated with chip types (Fine Chip and Good Chip) and the amount of bark present in the system, further supporting the PFMEA results that these variables substantially contribute to process risks. However, conventional statistical techniques, such as Pearson correlation and multiple linear regression, demonstrated limitations in capturing non-linear relationships and managing complex, imbalanced data distributions. As a result, machine learning models were introduced to improve predictive performance. Four models—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Neural Network (NN), and an Ensemble Method—were developed with appropriate data standardization and hyperparameter tuning prior to training. The comparative results indicated that no single model consistently outperformed others across all target variables. Nevertheless, the Neural Network achieved the highest accuracy in predicting occurrence (81.7%) and maintained consistent performance across all indicators, while the Ensemble Method attained the best accuracy in predicting detection (85.0%) and showed strength in handling imbalanced class distributions. Although the KNN model performed best in predicting severity (85.8%), its overall accuracy was lower in other variables, suggesting limited generalizability. Meanwhile, the SVM model offered stable results but achieved the lowest overall accuracy. Taken together, these findings indicate that both the Neural Network and Ensemble Method hold the most significant potential for future application in predictive risk analysis.

Compared with previous FMEA studies, this research demonstrates clear advantages. Keskin and Özkan [5] proposed the fuzzy ART algorithm as an alternative to traditional FMEA scoring; however, they did not test it with real-time production data. Balaraju et al. [6] applied a fuzzy-FMEA risk evaluation to a Load–Haul–Dump (LHD) machine in the mining sector, which performed well but still relied heavily on expert judgment. Spreafico and Sutrisno [7] developed a Social FMEA assisted by artificial intelligence to support sustainable product design, extending FMEA to social and environmental issues. Góes et al. [8] proposed a hybrid AI-based risk assessment framework combining ANN, fuzzy logic, and IoT for sustainable construction. In contrast, the present study integrates PFMEA with several machine learning algorithms directly on real MDF production data. It identifies key risk factors, such as the chipping process with an RPN of 192, and shows that the Neural Network and Ensemble Method achieve higher predictive accuracy. This integrated approach not only reduces subjectivity but also provides a practical framework that is ready for future integration with IoT sensors and transfer learning to enhance industrial applications.

Moreover, process improvement strategies can be guided by these findings. Priority should be given to controlling the proportions of fine chips and moisture, as identified in the Pareto analysis, while also managing pH levels during chip washing and minimizing blade wear. These objectives can be achieved by optimizing cutting speed and raw material feeding rates, implementing automated moisture control systems, continuously monitoring pH values, and replacing visual inspection with sensor-based detection. Integrating statistical process control (SPC) with machine learning—particularly through a web application for real-time monitoring and alerts—offers an effective strategy for enhancing predictive accuracy. Finally, increasing the volume of training data, balancing class distributions, and refining model parameters are essential steps to support more reliable and scalable risk management solutions in MDF production.

4. Conclusions

This study aimed to assess and predict risks in the medium-density fiberboard (MDF) production process by integrating Failure Mode and Effects Analysis (FMEA) with data analytics and machine learning techniques. The PFMEA effectively identified key risk factors in the pre-Refiner stages, particularly those associated with contamination and improper moisture levels, with the chipping process showing the highest Risk Priority Number (RPN) at 192. Statistical analysis confirmed that moisture content was significantly correlated with blade wear, specific energy consumption (SEC), chip types, and bark contamination. However, traditional methods, such as Pearson correlation and multiple linear regression, were inadequate for capturing complex, non-linear patterns and managing class imbalance. Four machine learning models—the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Neural Network (NN), and an Ensemble Method—were therefore developed and tested with data standardization and hyperparameter tuning. The KNN model achieved the highest accuracy in predicting severity (S), while the Neural Network performed strongly for occurrence (O). The Ensemble Method excelled in detection (D) and handled imbalanced data more robustly. In addition to model performance, this approach enables factories to adjust process parameters more precisely, reduce defects and waste, lower production costs, and minimize downtime, while providing a data-driven basis for engineering decisions and continuous improvement. The findings suggest that implementing these recommendations could reduce defect rates, improve process stability and yield, and lower maintenance costs. In addition, future work will focus on enhancing the model's performance and practical applicability. First, additional data will be collected through data augmentation to increase the diversity and representativeness of the training set. Second, the system will be connected to the Internet of Things (IoT) infrastructure, enabling real-time data acquisition from sensors installed at each stage of the production process, ranging from the debarking stage to the refinery. This real-time data stream enables continuous model updating and monitoring, thereby reducing latency between data collection and risk prediction. Finally, transfer learning techniques will be explored to leverage pre-trained models for faster adaptation and improved accuracy on new datasets.

5. Acknowledgments

The authors gratefully acknowledge the financial support provided by Thailand Science Research and Innovation (TSRI) under the Fundamental Fund (FF), fiscal year 2024, through the government budget of Thailand (Project Reference Code: 24463; Research Project Code: ENG6701262M). The authors would also like to extend their appreciation to the Smart Industrial Research Center, the Faculty of Engineering, and Prince of Songkla University for their support and provision of research resources and facilities.

Author Contributions: Conceptualization, Sirirat Pungchompoo and Nikorn Sirivongpaisal; Methodology, Sirirat Pungchompoo; Validation, Sirirat Pungchompoo; Formal analysis, Sirirat Pungchompoo; Investigation, Sirirat Pungchompoo, Nikorn Sirivongpaisal, Rakkrit Duansoithong, and Aree Teeraparbseree; Resources, Sirirat Pungchompoo; Data curation, Sirirat Pungchompoo, Nikorn Sirivongpaisal, Rakkrit Duansoithong, and Aree Teeraparbseree; Writing – original draft preparation, Sirirat Pungchompoo; Writing – review and editing, Sirirat Pungchompoo; Visualization, Sirirat Pungchompoo; Supervision, Sirirat Pungchompoo and Nikorn Sirivongpaisal; Project administration, Sirirat Pungchompoo and Nikorn Sirivongpaisal; Funding acquisition, Sirirat Pungchompoo. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Thailand Science Research and Innovation (TSRI) under the Fundamental Fund (FF), fiscal year 2024, grant number ENG6701262M. The APC was funded by Thailand Science Research and Innovation (TSRI).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

[1] Jiang, S.; Liu, Z.; Chen, J. A Dynamic Failure Mode and Effect Analysis (FMEA) Method for CNC Machine Tool in Service. *J. Phys.: Conf. Ser.* **2023**, *2483*, 012047. https://doi.org/10.1088/1742-6596/2483/1/012047

[2] Bangun, C. S.; Maulana, A.; Rasjidin, R.; Rahman, T. Application of SPC and FMEA Methods to Reduce the Level of Hollow Product Defects. *J. Tek. Ind.* **2022**, *8*(1), 12–16. https://doi.org/10.24014/jti.v8i1.16681

[3] Nurdaningsih, N. W.; Yunitasari, E. W.; Ma'arif, S. Statistical Process Control (SPC) and Fuzzy-Failure Mode and Effect Analysis (F-FMEA) Approaches to Reduce Reject Products in Wine Bottle Rack Production Process at PT Alis Jaya Ciptatama. *Opsi* **2022**, *15*(2), 274–283. https://doi.org/10.31315/opsi.v15i2.7567

[4] Appollis, L.-L. M.; van Dyk, W. A.; Matope, S. Using Failure Modes and Effects Analysis as a Problem-Solving Guideline When Implementing SPC in a South African Chemical Manufacturing Company. *S. Afr. J. Ind. Eng.* **2020**, *31*(1), 157–169. https://doi.org/10.7166/31-1-2294

[5] Keskin, G. A.; Özkan, C. An Alternative Evaluation of FMEA: Fuzzy ART Algorithm. *Qual. Reliab. Eng. Int.* **2009**, *25* (6), 647–661. https://doi.org/10.1002/qre.984

[6] Balaraju, J.; Govinda Raj, M.; Murthy, C. S. Fuzzy-FMEA Risk Evaluation Approach for LHD Machine – A Case Study. *J. Sustain. Min.* **2019**, *18*(4), 257–268. https://doi.org/10.1016/j.jsm.2019.08.002

[7] Spreafico, C.; Sutrisno, A. Artificial Intelligence Assisted Social Failure Mode and Effect Analysis (FMEA) for Sustainable Product Design. *Sustainability* **2023**, *15*(11), 8678. https://doi.org/10.3390/su15118678

[8] Góes, A. L. B.; Kazmi, R.; Aqsa, A.; Nuthakki, S. A Hybrid AI-Based Risk Assessment Framework for Sustainable Construction: Integrating ANN, Fuzzy Logic, and IoT. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*(3), 46–56.

[9] Lu, Q.; Zeng, W.; Guo, Q.; Lü, S. Optimal Operation Scheduling of Household Energy Hub: A Multi-Objective Optimization Model Considering Integrated Demand Response. *Energy Rep.* **2022**, *8*, 15173–15188. https://doi.org/10.1016/j.egyr.2022.11.047

[10] Doz, D.; Cotič, M.; Felda, D. Random Forest Regression in Predicting Students' Achievements and Fuzzy Grades. *Mathematics* **2023**, *11*(19), 4129. https://doi.org/10.3390/math11194129

[11] Muchlinski, D.; Siroky, D.; He, J.; Kocher, M. Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data. *Political Analysis* **2016**, *24*(1), 87–103. https://doi.org /10.1093/pan/mpv024

[12] Program, D. F.; Prijadi, R. FMEA-Based Logistic Regression Model for the Evaluation of Photovoltaic Power Plant Risk. *Quant. Econ. Manag. Stud.* **2024**, *5*, 644-657. https://doi.org/10.35877/454RI.qems2645