Predicting Breakdown Types of Biomedical Equipment in Public Hospital Using Machine Learning

M.Luqman Zulkepli[1, a)] and Suryani Ilias[1, b)]

[1] *Jabatan Kejuruteraan Elektrik, Politeknik Sultan Salahuddin Abdul Aziz Shah, Selangor, Malaysia*

[a)]Corresponding author: mluqmanzul@gmail.com
[b)]kamasue77@gmail.com

Received 1 October 2025, Accepted 28 October 2025, Available Online 30 November 2025

Abstract. Predictive maintenance has emerged as a vital methodology in biomedical engineering to ensure the reliability of life-support equipment, particularly ventilators. This study presents a machine learning-based approach for predicting ventilator breakdown types using unstructured historical maintenance data. The dataset, consisting of 10,314 unscheduled maintenance records, was obtained from the Asset and Services Information System (ASIS), managed by the Ministry of Health Malaysia (KKM). The primary objective is to classify technician-written, free-text fault descriptions into structured breakdown categories to enable proactive maintenance decision-making. Four machine learning models, Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and Random Forest (RF) were developed and evaluated under two workflows: one using basic label encoding, and the other incorporating Term Frequency–Inverse Document Frequency (TF-IDF) for semantic feature extraction. The results demonstrate that TF-IDF significantly enhanced classification performance across all models. SVM achieved the highest accuracy of 94.70%, followed by DT (94.56%), RF (94.14%), and ANN (92.19%). In contrast, the same models trained without TF-IDF performed poorly, with accuracies falling below 36% for SVM and ANN. These findings emphasize the importance of textual feature engineering in predictive maintenance applications and validate the effectiveness of AI-based text classification for unstructured biomedical data. By operationalizing real-world maintenance notes from clinical engineering workflows, this study demonstrates the practical viability of machine learning in improving equipment uptime, optimizing service scheduling, and enhancing patient safety in healthcare settings.

Keywords: predictive maintenance, TF-IDF, ventilator, machine learning, classification.

INTRODUCTION

Modern healthcare systems heavily depend on the proper functioning of biomedical equipment to deliver safe and efficient patient care. In high-stakes environments like Intensive Care Units (ICUs), ventilators serve as vital lifesupport systems for patients experiencing respiratory failure or undergoing surgical procedures.. Studies show that equipment-related issues contribute to as much as 81% of reported safety incidents, with ventilators representing a major concern when maintenance procedures are lacking (1). Conventional maintenance methods are often reactive, lacking the ability to foresee potential failures, which leads to decreased performance and prolonged downtime, particularly problematic for critical medical devices (2). Evidence has linked ventilator malfunctions to negative patient outcomes, underlining the urgent need for more dependable maintenance solutions (3).

Ventilators play a vital role for patients undergoing surgical procedures or suffering from acute respiratory distress. Given their intricate design which includes components like compressors, sensors, valves, and control software, rigorous maintenance is essential to guarantee safe operation. Even minor technical issues can lead to life-threatening consequences, making predictive maintenance strategies vital for risk prevention (4). Although ventilators are critical for saving lives, they inherently carry certain risks. This makes implementing robust risk management practices in their use and upkeep indispensable in intensive care environments. Ongoing monitoring and consistent care are required to maintain effective ventilation (5). Because of their complex mechanical, electrical, and software elements, regular maintenance is necessary to prevent serious complications such as ventilator-induced lung injury (VILI) and hospital-acquired infections. As a result, adopting intelligent, data-driven maintenance systems is not just costefficient but also fundamental for ensuring patient safety.

Healthcare maintenance practices generally encompass corrective, preventive, condition-based, and predictive maintenance (PdM). Predictive maintenance utilizes machine learning algorithms to forecast potential equipment failures before they happen. The healthcare industry is increasingly embracing predictive maintenance (PdM) methods, driven by advancements in real-time analytics and the integration of Internet of Things (IoT) technologies (6). The convergence of these innovations enables seamless data acquisition from biomedical equipment, allowing for continuous performance monitoring and early fault detection. Real-time analytics processes this data to uncover actionable insights, while IoT ensures timely and automated data transmission from devices in clinical environments. This forward-thinking strategy not only reduces unnecessary preventive maintenance tasks but also helps minimize unplanned equipment downtime, thereby enhancing operational efficiency and patient safety (7). Notable attributes of PdM include real-time remote monitoring and advanced diagnostic tools that aid in early fault detection, thereby avoiding expensive repairs. This approach contributes to a reliable, cost-effective, and environmentally sustainable maintenance model that ensures consistent equipment performance (8).

The integration of Internet of Things (IoT)-based monitoring has significantly advanced continuous data collection on biomedical equipment performance. Artificial Intelligence (AI) algorithms process this data to forecast potential failures and initiate timely maintenance actions (9). Specifically, an IoT-enabled predictive maintenance framework for ICU ventilators employs real-time monitoring and data analytics to anticipate malfunctions, with a focus on highrisk components such as the pneumatic block identified through Failure Modes and Effects Analysis (FMEA). This strategy not only increases system reliability and operational efficiency but also minimizes patient risk in critical care environments (6).

Recent studies have demonstrated that classification algorithms like Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees (DT), and Random Forests (RF) are highly effective for predictive maintenance tasks involving both textual and sensor data. SVM is known for its strong generalization ability, making it especially effective in high-dimensional spaces, such as those formed by Term Frequency–Inverse Document Frequency (TF-IDF) vectors (10). ANN models, in contrast, excel at identifying complex nonlinear patterns within noisy biomedical maintenance data, which enhances classification accuracy and allows the system to detect intricate failure signatures (11).

DT models are prized for their interpretability and straightforward logic-based branching structure, making them suitable for initial deployment in resource-constrained healthcare settings. However, they are often limited by their tendency to overfit. RF addresses this limitation by constructing an ensemble of multiple decision trees using bootstrap aggregation and random feature selection, which significantly improves robustness, reduces variance, and enhances classification performance. In recent applications, RF has demonstrated high reliability and fault detection accuracy in classifying diverse biomedical equipment failures, offering a scalable and interpretable solution for predictive maintenance in clinical environments (12). These machine learning models collectively support the development of early warning mechanisms that reduce risk, minimize equipment downtime, and improve healthcare service continuity.

This study aims to automate the classification of unstructured maintenance records into structured breakdown categories using machine learning algorithms, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and Random Forest (RF). It also examines the effectiveness of feature extraction techniques, particularly Term Frequency–Inverse Document Frequency (TF-IDF), in enhancing classification accuracy. The objective is to predict breakdown types from historical maintenance data, enabling timely decisionmaking and proactive maintenance actions to improve equipment reliability and operational efficiency.

Additionally, this paper is divided into four sections. Section II discusses the methodology used, Section III presents the results and provides the discussion and analysis, and Section IV concludes the findings

RESEARCH METHOD

This study employed maintenance records obtained from the Asset and Services Information System (ASIS), specifically under the category "Service_Work_(Unscheduled_&_Others)," collected from three public hospitals spanning the period from December 2001 to December 2024. The initial dataset comprised 10,314 entries, each documenting instances of unscheduled maintenance. To enable machine learning analysis, a comprehensive data preprocessing pipeline was developed using Python, incorporating libraries such as Pandas, NumPy, Scikit-learn, and FuzzyWuzzy for data cleaning, transformation, and standardization of unstructured text.

Initially, entries marked as "Cancelled" in the Work Order Status column were excluded to ensure the dataset comprised only valid maintenance records. The core of the analysis centred on the Response Finding column, which contains textual descriptions of identified faults or issues. Records lacking values in this column were also removed to ensure that only entries with relevant diagnostic information were retained for further processing.

Subsequently, the Response Finding column underwent a sequence of text preprocessing steps. These included converting all text to lowercase, removing punctuation marks, and eliminating irrelevant or non-informative characters to standardize the data. To further refine the dataset, a domain-specific fuzzy matching algorithm was applied using the FuzzyWuzzy library, following a rule-based mapping system informed by expert knowledge (13). This process enabled the transformation of unstructured textual maintenance descriptions into structured, analysable labels. As a result of this approach, the dataset was successfully classified into seven predefined ventilator breakdown categories: Power Issue, Oxygen Sensor Issue, Technical Issue, Structural Issue, Control Interface Issue, Flow Sensor Issue, and Electronics Issue, as summarized in TABLE 1.

To ensure a clean and reliable dataset for training machine learning models, unclassified entries that did not align with any of the established rule-based categories were removed. This step helped maintain data consistency and relevance for downstream analysis. For converting textual data into a numerical format compatible with machine learning algorithms, two different approaches were implemented. The first approach involved label encoding without feature extraction. Using Scikit-learn's LabelEncoder, each unique string in the Response Finding column was mapped to a distinct integer. While this method does not preserve semantic meaning or context, it served as a baseline to satisfy the numerical input requirement for classification models.

The second approach employed a more advanced technique, Term Frequency–Inverse Document Frequency (TFIDF) to capture semantic relationships in the text. Implemented via Scikit-learn's TfidfVectorizer, this method was configured with an n-gram range of (1–3), a maximum document frequency of 85%, and a minimum document frequency of 2. These parameters were chosen to emphasize informative words and multi-word expressions while excluding overly common or rare terms. The result was a high-dimensional sparse matrix representing the textual data, offering greater semantic richness and improved suitability for classification tasks. TF-IDF was chosen for its welldocumented success in improving model performance when applied to unstructured textual inputs, especially in combination with machine learning classifiers (13).

TABLE 1. Description of classified ventilator breakdown types

No	Breakdown Types	Description
1	Power Issue	Failures related to electrical supply, battery, or voltage regulation affecting ventilator operation.
2	Oxygen Sensor Issue	Malfunctions in oxygen measurement, regulation, or sensor failures impacting oxygen delivery accuracy.
3	Technical Issue	Software, firmware, or system-related errors, including calibration failures, alarms, and diagnostic issues.
4	Structural Issue	Physical damage or integrity failures in ventilator casing, mounts, tubing, or chassis.
5	Control Interface Issue	Problems with user interaction components such as screens, buttons, keypads, or control panels.
6	Flow Sensor Issue	Failures in breath detection, expiratory circuits, or airflow measurements affecting ventilation performance.
7	Electronics Issue	Internal electronic component failures, including PCB, temperature control, relay, or circuit malfunctions.

This dual approach enabled a comparative evaluation of simple label encoding versus semantic feature extraction using TF-IDF and their respective impacts on classification performance. Four machine learning classifiers were developed and tested using Scikit-learn: Support Vector Machine (SVM), Artificial Neural

Network (ANN), Decision Tree (DT), and Random Forest (RF). These models were chosen based on their proven effectiveness in predictive maintenance tasks. SVM is noted for its strong performance in high-dimensional spaces created by textual data. ANN excels at identifying complex, nonlinear patterns often found in maintenance logs. DT offers high interpretability, while RF enhances robustness and accuracy by creating an ensemble of decision trees. The labelled dataset was divided into training and testing subsets using an 80:20 stratified sampling technique, ensuring each of the seven

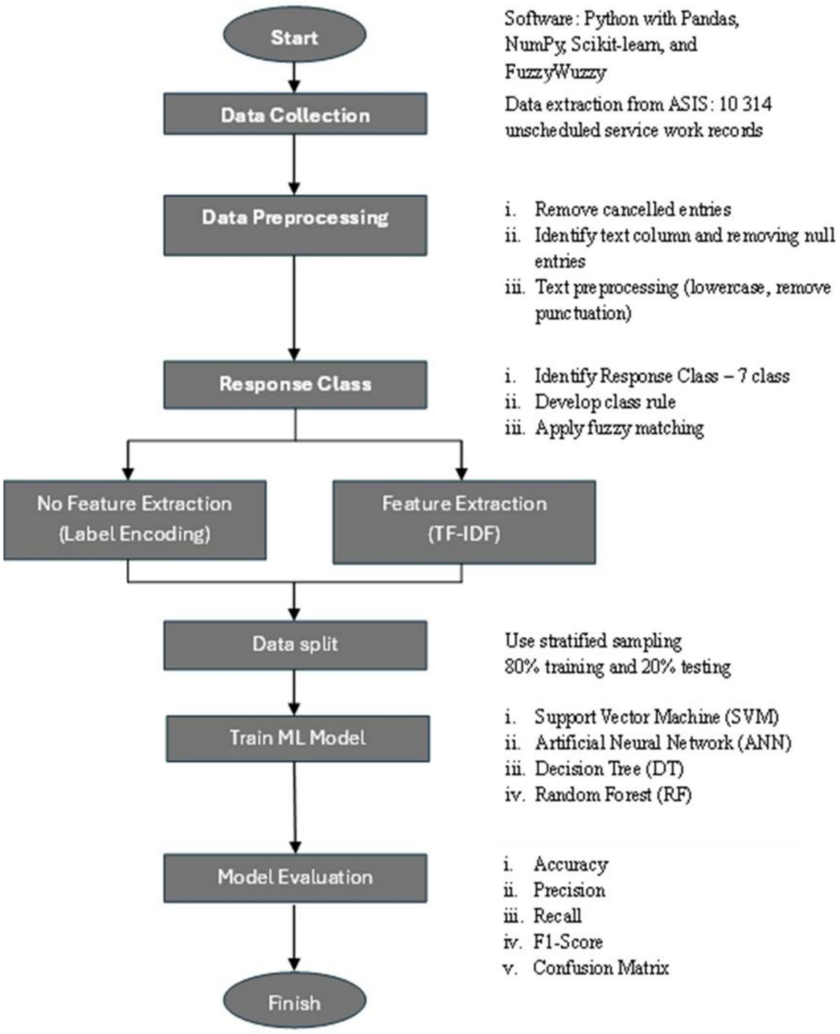

FIGURE Workflow of the predictive maintenance methodology.

breakdown categories was proportionally represented in both sets. This approach promoted balanced training and improved the validity of performance assessments.

Model performance was evaluated using four key metrics: accuracy, precision, recall, and F1-score, as recommended by Naidu et al. (2023). Additionally, a confusion matrix was generated for each model to visualize the distribution of correctly and incorrectly classified instances across all breakdown categories, providing deeper insights into classification effectiveness (15). The formulas for the evaluation metrics are defined as follows: Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$, Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$ and F1-Score = $2 \times \frac{Precision \times Recall}{Precision + Rec}$,where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives. Additionally, two types of confusion matrices were used. The absolute confusion matrix displays the raw counts of true versus predicted labels for each class, represented as:

$$Confusion\ Matrix\ (Absolute) = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{12} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

Where c_{ij} represents the number of instances where class i was predicted as class j. The normalized confusion matrix displays the proportion of predictions for each true class, calculated as:

$$Confusion\ Matrix\ (Normalized) = \frac{c_{ij}}{\sum_{j=1}^{n} c_{ij}}\ for\ each\ row\ i$$

This normalized version ensures each row sums to 1, allowing for per-class performance comparison regardless of sample size. A complete overview of this methodology is illustrated in **FIGURE 1.**

RESULTS AND DISCUSSION

TABLE 2. Performance Comparison of Machine Learning Models with and without Feature Extraction

Model	Feature Extraction	Accuracy	Precision	Recall	F1 Score
Random Forest	Without TF-IDF	77.96%	78.48%	77.96%	78.14%
	With TF-IDF	94.14%	94.32%	94.14%	94.12%
Decision Tree	Without TF-IDF	77.96%	78.48%	77.96%	78.14%
	With TF-IDF	94.56%	94.58%	94.56%	94.56%
SVM	Without TF-IDF	33.61%	15.97%	33.61%	21.03%
	With TF-IDF	94.70%	94.73%	94.70%	94.68%
ANN	Without TF-IDF	35.70%	20.22%	35.70%	23.06%
	With TF-IDF	92.19%	92.15%	92.19%	92.15%

TABLE 2. presents a comparative analysis of the performance of four machine learning models, Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM) and Artificial Neural Network (ANN) with and without the application of TF-IDF feature extraction. The models were evaluated using four standard classification metrics: accuracy, precision, recall, and F1-score. When TF-IDF vectorization was applied, all models exhibited substantial performance improvements across every metric. SVM achieved the highest performance overall, with an accuracy of 94.70% and an F1-score of 94.68%, demonstrating excellent predictive capability and generalization. ANN also performed well, with an accuracy of 92.19% and an F1-score of 92.15%, highlighting its ability to capture nonlinear patterns in semantically enriched text. Notably, DT and RF models, which previously had limited expressiveness when using raw label-encoded text, reached 94.56% and 94.14% accuracy, respectively, after applying TF-IDF. Both achieved F1-scores above 94%, confirming their improved classification capability with semantic features.

In contrast, without TF-IDF, model performance declined sharply. DT and RF each yielded 77.96% accuracy and 78.14% F1-scores, adequate, but notably lower than with semantic enhancement. SVM and ANN performed the worst under the non-feature-extracted condition, with SVM achieving only 33.61% accuracy and an F1-score of 21.03%, while ANN recorded 35.70% accuracy and an F1-score of 23.06%. These results shows that raw label encoding fails to provide meaningful representation for unstructured maintenance text, underscoring the necessity of effective feature extraction like TF-IDF in predictive maintenance applications.

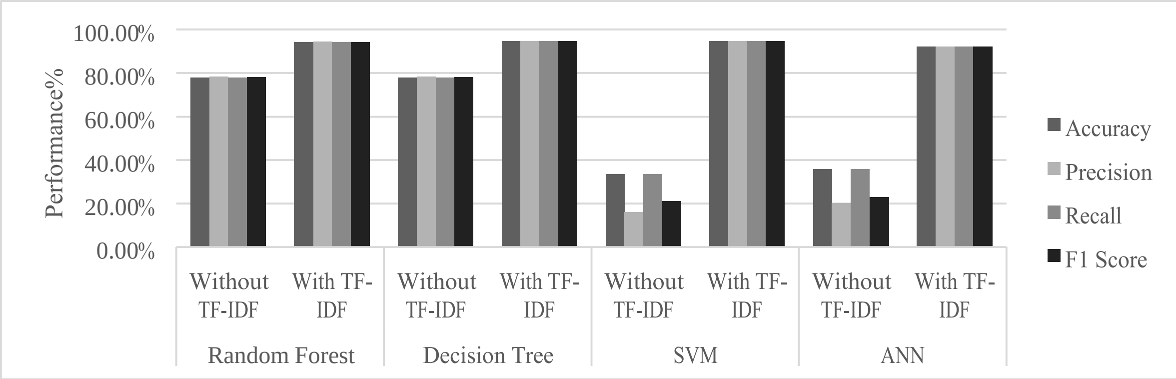

FIGURE 2. Comparative performance of machine learning models (Random Forest, Decision Tree, Support Vector Machine, and Artificial Neural Network) with and without TF-IDF feature extraction.

FIGURE 2. provides a visual comparison of the performance metrics, accuracy, precision, recall, and F1-score across four machine learning models (SVM, ANN, Decision Tree, and Random Forest), evaluated both with and without TF-IDF feature extraction. The chart demonstrates a clear and consistent performance boost across all models when TF-IDF is applied. The improvement is especially striking for SVM and ANN, which initially performed poorly with raw, label-encoded text. Upon applying TF-IDF, their accuracy and F1-scores more than doubled, confirming the importance of semantic feature extraction in handling unstructured maintenance data.

While DT and RF already delivered moderate performance without TF-IDF, their results significantly improved after applying feature extraction, both models reaching F1-scores above 94%. This supports the finding that tree-based classifiers, when combined with high-quality textual features, can be highly effective for classification tasks in biomedical maintenance contexts. Overall, Figure 2 underscores the critical role of textual preprocessing and feature engineering, particularly TF-IDF in enhancing the predictive performance of machine learning models used in preventive maintenance frameworks for healthcare equipment.

'FIGURE 3.' presents the normalized confusion matrices for four machine learning models, Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Artificial Neural Network (ANN) trained using TF-IDF feature extraction to classify ventilator breakdown types across seven predefined categories. The normalization ensures that each row sums to 1, allowing for direct interpretation of classification accuracy per class, independent of class imbalance. Higher diagonal values indicate more accurate predictions for each specific breakdown category.

- Random Forest with TF-IDF (Top-Left): Demonstrates robust classification performance, achieving high recall for Technical Issue (97%), Flow Sensor Issue (94%), and Oxygen Sensor Issue (91%). Slight confusion is observed in predicting Electronics (83%) and Structural Issues (87%), likely due to feature overlap in the maintenance text. Overall, RF offers a balanced performance across all categories.
- Support Vector Machine with TF-IDF (Top-Right): Outperforms all other models with consistently high accuracy across categories. It correctly classifies 99% of Oxygen Sensor Issues, 96% of Flow Sensor Issues, and 94% of Technical Issues. This matrix highlights SVM's strength in handling high-dimensional feature spaces created by TF-IDF, making it well-suited for distinguishing between nuanced textual patterns in unstructured data.

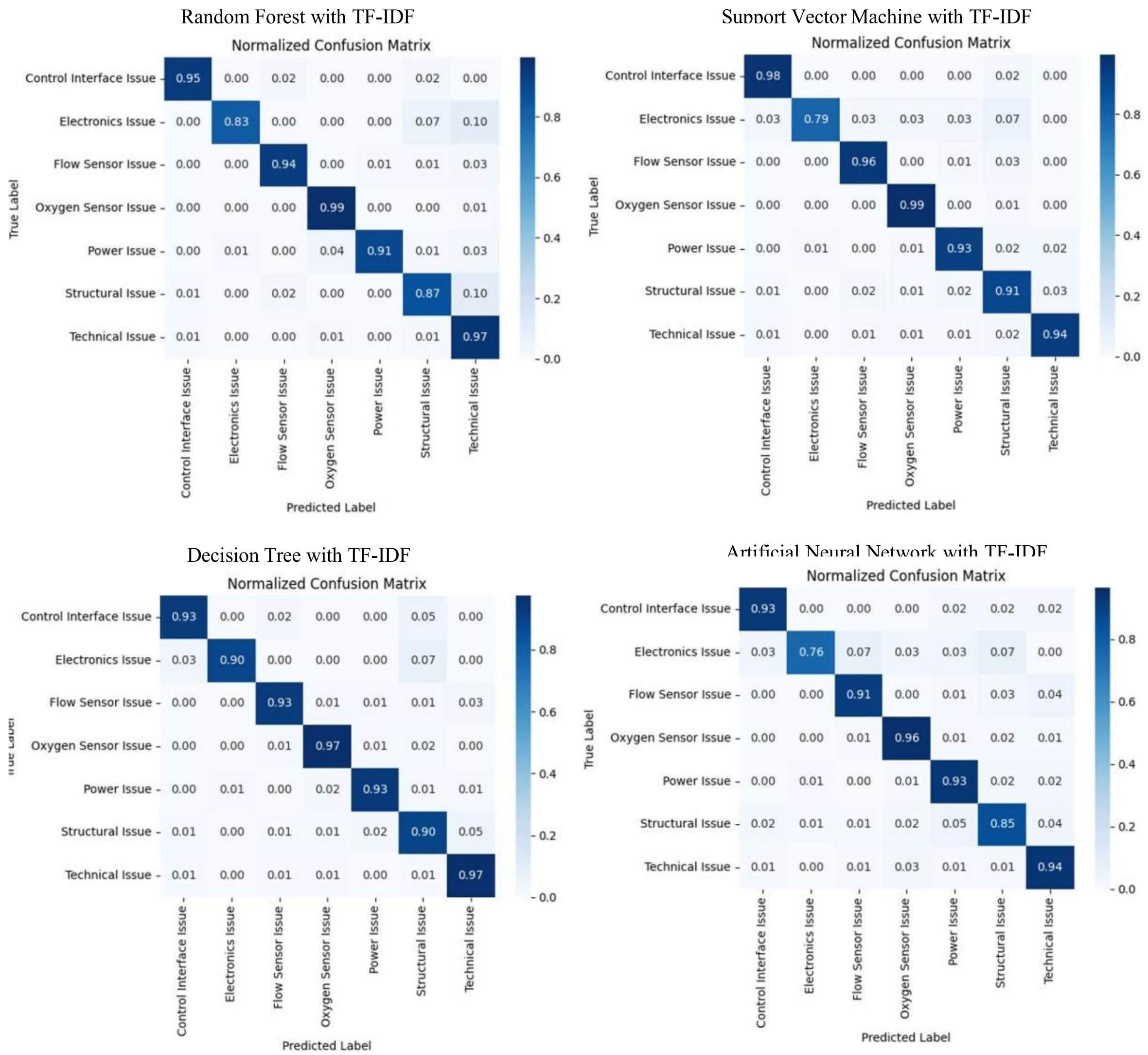

FIGURE 3. Normalized confusion matrices for Random Forest, Support Vector Machine, Decision Tree, and

Artificial Neural Network models trained with TF-IDF feature extraction. Each matrix displays the per-class classification performance across seven ventilator breakdown categories. Higher diagonal values indicate stronger class-specific accuracy.

- Decision Tree with TF-IDF (Bottom-Left): Shows slightly better performance than RF in certain categories, with strong predictions for Technical Issue (97%) and Flow Sensor Issue (93%). Notably, it achieves 90% recall for both Electronics Issue and Structural Issue, outperforming RF in these categories. This suggests that DT may have better memorization of rule-based splits specific to these fault types, though it also carries a higher risk of overfitting.

- Artificial Neural Network with TF-IDF (Bottom-Right): Performs competitively, particularly in predicting Technical Issue (97%) and Power Issue (93%). However, it records the lowest recall for Electronics Issue (76%) and Structural Issue (85%) among the four models. The misclassification of 'Electronics Issue' and 'Structural Issue' may stem from overlapping terminology in the free-text descriptions. For instance, a description of a faulty circuit board ('Electronics Issue') might share keywords with general system errors ('Technical Issue'), making it difficult for the ANN to discern the unique failure pattern without more advanced feature engineering.

In summary, the confusion matrices affirm that TF-IDF significantly enhances classification performance across all models. Among them, SVM delivers the most reliable and consistent results, making it a preferred choice for predictive maintenance systems that rely on unstructured text data.

CONCLUSION

This study demonstrates the effectiveness of machine learning algorithms Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and Random Forest (RF), in classifying unstructured maintenance records of biomedical equipment into structured breakdown categories. Using historical data from the Asset and Services Information System (ASIS) and applying TF-IDF feature extraction, all four models showed notable improvements in predictive accuracy, precision, recall, and F1-score compared to baseline models without feature extraction.

Among the tested models, SVM with TF-IDF achieved the highest performance, with an accuracy of 94.70% and an F1-score of 94.68%, followed closely by DT (94.56%), RF (94.14%), and ANN (92.19%). These results highlight the importance of semantic feature representation when dealing with unstructured biomedical maintenance data. In contrast, models trained without TF-IDF underperformed significantly especially SVM and ANN illustrating that raw label-encoded text lacks the contextual depth needed for accurate classification.

The use of normalized confusion matrices further provided insight into model behaviour at the class level. SVM and DT demonstrated strong consistency across all breakdown categories, while RF and ANN performed slightly lower in cases where textual descriptions were ambiguous or overlapping. Despite these differences, all TF-IDFenhanced models proved effective in identifying critical faults such as Oxygen Sensor Issues, Technical Issues, and Power Failures.

This research also operationalizes the recommendation of Rahman et al. (2023), who emphasized integrating unstructured maintenance notes for predictive insights. By utilizing technician-written descriptions from the Response Finding field typically recorded immediately after service, the study enhances the contextual relevance of the input data and mirrors real-world clinical engineering practices.

In conclusion, the implementation of AI-driven predictive maintenance using machine learning models, particularly with TF-IDF feature extraction, offers a practical and scalable solution for enhancing biomedical equipment reliability. This is especially critical for ventilators, where early fault prediction can directly impact patient outcomes in intensive care settings. The findings support further adoption of intelligent maintenance systems in healthcare to reduce equipment downtime, improve service efficiency, and strengthen patient safety.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Engineering Division of the Ministry of Health Malaysia for providing access to the maintenance data from the Asset and Services Information System (ASIS), whose support and cooperation were instrumental in the successful execution of this research project. The authors also extend their heartfelt appreciation to Associate Professor Dr. Ahmad Ihsan Mohd Yassin, Associate Professor at the Microwave Research Institute, Universiti Teknologi MARA (UiTM), for his invaluable guidance, expert insights, and encouragement throughout the course of this study.

REFERENCES

Ratwani R, Adams K, Kim T, Busog DN, Howe J, Jones R, et al. Assessing Equipment, Supplies, and Devices for Patient Safety Issues. Patient Safety. 2023 Mar 31;5(1):15–25.

Manchadi O, Ben-Bouazza FE, Jioudi B. Predictive Maintenance in Healthcare System: A Survey. IEEE Access. 2023;11:61313–30.

Yoshioka J, Nakane M, Kawamae K. Healthcare Technology Management (HTM) of mechanical ventilators by clinical engineers. Vol. 2, Journal of Intensive Care. BioMed Central Ltd.; 2014.

Lei Y. Medical Ventilator System Basics: A Clinical Guide. Oxford University Press; 2017.

Arathy M, Karthi Balasubramaniam. Fuzzy Logic based Failure Modes and Effects Analysis on Medical Ventilators. In: Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020). IEEE; 2020.

Manchadi O, El Otmani Dehbi Z, Ben-Bouazza FE, Edder A, Tafala I, Jioudi B. IoT-Powered Predictive Maintenance Framework for ICU Ventilators. In: 2024 International Conference on Intelligent Systems and Computer Vision, ISCV 2024. Institute of Electrical and Electronics Engineers Inc.; 2024.

Abdul Jamil AS, Khalil A, Mohd Yunus M, Fofah JG. Empowering Predictive Maintenance of Medical Equipment Through AI-Driven Condition Monitoring. 2024 [cited 2025 Apr 17];53–68. Available from: https://link.springer.com/chapter/10.1007/978-981-97-9294-8_4

Achouch M, Dimitrova M, Ziane K, Sattarpanah Karganroudi S, Dhouib R, Ibrahim H, et al. On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges. Vol. 12, Applied Sciences (Switzerland). MDPI; 2022.

Boppana V. Data Analytics for Predictive Maintenance in Healthcare Equipment. EPH - International Journal of Business & Management Science. 2023 May;9(2):78–87.

Zonta T, da Costa CA, da Rosa Righi R, de Lima MJ, da Trindade ES, Li GP. Predictive maintenance in the Industry 4.0: A systematic literature review. Comput Ind Eng. 2020 Dec 1;150:106889.

Molina E, Parraga-Alava J. Artificial Neural Networks for Classification Tasks: A Systematic Literature Review. Enfoque UTE [Internet]. 2024 Oct 1;15(4):1–10. Available from: https://ingenieria.ute.edu.ec/enfoqueute/index.php/revista/article/view/1058

Argente-Garrido A, Zuheros C, Luzón MV, Herrera F, Argente-Garrido A, Zuheros C, et al. AN INTERPRETABLE CLIENT DECISION TREE AGGREGATION PROCESS FOR FEDERATED LEARNING. 2024;

Bounabi M, El Moutaouakil K, Satori K. Text classification using Fuzzy TF-IDF and Machine Learning Models. In: ACM International Conference Proceeding Series. Association for Computing Machinery; 2019.

Naidu G, Zuva T, Sibanda EM. A Review of Evaluation Metrics in Machine Learning Algorithms. Lecture Notes in Networks and Systems [Internet]. 2023 [cited 2025 Apr 17];724 LNNS:15–25. Available from: https://link.springer.com/chapter/10.1007/978-3-031-35314-7_2

Sathyanarayanan S, Tantri BR. Confusion Matrix-Based Performance Evaluation Metrics. African Journal of Biomedical Research [Internet]. 2024 Nov 30 [cited 2025 Apr 17];27(4S):4023–31. Available from: https://mail.africanjournalofbiomedicalresearch.com/index.php/AJBR/article/view/4345

Rahman NHA, Zaki MHM, Hasikin K, Razak NAA, Ibrahim AK, Lai KW. Predicting medical device failure: a promise to reduce healthcare facilities cost through smart healthcare management. PeerJ Comput Sci. 2023;9:1–34.