IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 12 August 2023, accepted 24 August 2023, date of publication 29 August 2023, date of current version 5 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3309671

RESEARCH ARTICLE

Medical Device Failure Predictions Through AI-Driven Analysis of Multimodal Maintenance Records

NOORUL HUSNA ABD RAHMAN[1,2], KHAIRUNNISA HASIKIN[1,3], NASRUL ANUAR ABD RAZAK[1], AYMAN KHALLEL AL-ANI[4], D. JERLINE SHEEBHA ANNI[5], AND PRABU MOHANDAS[6]

[1]Department of Biomedical Engineering, University of Malaya, Lembah Pantai, Kuala Lumpur 50603, Malaysia
[2]Engineering Services Division, Ministry of Health, Putrajaya 62590, Malaysia
[3]Center of Intelligent Systems for Emerging Technology (CISET), Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia
[4]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Sabah 88400, Malaysia
[5]Department of Computer Science and Engineering, KMCT College of Engineering for Women, Kozhikode, Kerala 673601, India
[6]Intelligent Computing Lab, Department of Computer Science and Engineering, National Institute of Technology, Calicut, Kozhikode, Kerala 673601, India

Corresponding author: Khairunnisa Hasikin (khairunnisa@um.edu.my)

This work was supported in part by the Ministry of Higher Education through MRUN Young Researchers Grant Scheme (MY-RGS) under Grant MR001-2019, in part by Universiti Malaya Living Laboratory under Grant LL2022CN002, and in part by Universiti Malaya Grant under Grant RMF0401-2021.

ABSTRACT Medical device failure and maintenance records are essential information, as some nations lack dedicated systems for capturing this valuable data. In addition to making healthcare more intelligent and individualized, machine learning has the potential to transform the conventional healthcare system. Optimizing AI models in decision-making could mitigate the effects of three research issues: malfunctioning medical devices, high maintenance costs, and the lack of a strategic maintenance framework. This study proposes a data-driven machine-learning model for predicting medical device failure. The proposed predictive model is developed using multimodal data of structured maintenance and unstructured text narrative of maintenance reports to predict the failure of 8,294 critical medical devices. In developing the model, 44 varieties of essential medical devices from 15 healthcare institutions in Malaysia are utilized. A classification problem is addressed by classifying failure into three prediction classes: (i) class 1, unlikely to fail within the first three years, (ii) class 2, likely to fail within three years; and (iii) class 3, likely to fail after three years from the date of commissioning. The topic modelling and synthesis strategy: Latent Dirichlet Allocation is applied to unstructured data in order to uncover concealed patterns in maintenance notes captured during failures. In addition, sensitivity analysis is performed to select only the most significant parameters affecting the failure performance of the medical device. Then, four machine learning algorithms and three deep learning networks are evaluated to determine the best predictive model. Based on the performance evaluation, the Ensemble Classifier is further optimized and demonstrates improved accuracy of 88.80%, specificity of 94.41%, recall of 88.82%, precision of 88.46%, and F1 Score of 88.84%. The study proves a reduction in intervention from 18 to 8 features and a reduction in training time from 1660.5 to 901.66 seconds for comprehensive model development.

INDEX TERMS Artificial intelligence, machine learning, medical device failure prediction, medical device maintenance, maintenance cost.

I. INTRODUCTION

The complexity of managing medical devices has increased in recent years due to the equipment's increasing sophistication and specialization, its integration with electronic networks,

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

and its reliance on outsourcing for specialized maintenance and repair. The internet of things application has evolved by using sensor installation in individual equipment to capture data. However, the discrete sensor and internet coverage to all facilities incurred a significant cost in facility management. In addition, the compliance, safety, and reliability criteria are always rising and undergoing rapid transformation [1].

The World Health Organization (WHO) describes pre-market, placing on-market and post-market as the three main stages in regulating medical device law. Malaysia, for example, has enacted Medical Device Act 2012 (Act 737) and Medical Device Authority Act 2012 (Act 738) and embarked on these two regulations in Medical Device Regulatory System. Act 737 regulates medical devices in the industry; meanwhile, Act 738 creates the Medical Device Authority for law enforcement. Any apparatus or instruments used, alone or in combination, for diagnosis, prevention, monitoring, treatment, or injury are considered medical devices [2].

The maintenance strategy aims to assure medical devices' dependability, maintainability, availability, and safety. A WHO reported that 50% to 80% of devices are dysfunctional due to the lack of maintenance culture and skills [3]. Numerous studies have found a correlation between catastrophic injuries and patient fatalities caused by malfunctioning medical devices [4], [5], [6], [7]. According to WHO estimation, over 50% of medical devices in low-middle-income countries are non-functional, not fully utilized and maintained [8].

By then, failure in user handling, use of refurbished or procurement of non-original equipment manufacturer's spare parts, random failure, incompetent repair technique, and wear-out failures are among the significant causes of equipment failures [3]. Besides, design defects and other environmental factors and stress, such as electromagnetic interference, high temperatures or humidity, may impact the device's functionality.

The market for maintaining medical devices increased tremendously, estimated at USD 35.3 billion in 2020, and it is anticipated to grow at a compound annual growth rate (CAGR) of 7.9 percent from 2021 to 2027. The sophisticated increase in medical device maintenance market is driven by increasing demand for medical devices and an increase in the market for refurbished medical devices [9]. The recent installation reported by Markets and Markets [10] denotes a significant increase in cost with the service contract, which demands 12% of the device's cost to be spent yearly. Throughout the device's lifecycle, the overall maintenance costs are more substantial than the device's cost [10], [11].

The government continues the burden of the high maintenance costs with ageing medical devices due to limited budget allocation for device replacement every year. The medical device replacement policy has to be strengthened, and the scenario worsens with unlimited maximum age for medical device usage. For example, in Malaysia, it is stated in Auditor's General Report published in 2021 that 19.6% of medical devices with more than 20 years of age and 11.5% have reached the end of life but are still in service [11]. The value of medical devices more than 20 years in age is equivalent to Malaysian Ringgit (MYR) 641.59 million to be replaced and is impractical to be executed simultaneously.

There are three main types of maintenance: corrective maintenance, preventive and predictive maintenance. Corrective maintenance is performed once a failure is discovered to return devices to their working state. In contrast, preventative maintenance is performed in a predefined interval as a preventive action before failure, lowering the risk of medical device breakdown or deterioration [1]. In order to minimize future failure, predictive maintenance is performed and is a proactive task for future failure predictions.

During the COVID-19 pandemic, the use of big data analytics to foresee the risk of failures has garnered much attention due to its predictive potential. In medical device maintenance, there are two types of maintenance data: structured and unstructured [12]. An example of structured data is the device age, while an example of unstructured data is failure notes recorded in the log history. The majority of extant literature employs numerical and categorical forms of structured data. To the best of our knowledge, none of the existing studies utilizes text narratives input in medical device performance prediction studies.

This research proposes an enhanced study by presenting multimodal data for the failure prediction of medical devices. An example unstructured data application for modelling is Latent Dirichlet Allocation (LDA). An LDA is a topic modelling approach and one of the most effective for text document analysis [13].

Concerning the research problems related to malfunctioning medical devices, high maintenance cost and ageing devices, a predictive model is proposed driven by the following objectives:

- To determine medical device failure occurrence patterns from multimodal unstructured maintenance reports.
- To integrate multimodal structured and unstructured text narrative maintenance reports for failure prediction.
- To propose a failure predictive model based on multimodal data using machine learning and deep learning techniques.
- To perform cost analysis based on a predictive medical devices failure model.

This study provides a failure prediction for three classes of medical devices. The article is organized as follows. Section II discusses the related works that apply medical devices, machine learning and deep learning studies. Section III explains the nature of the dataset, machine learning, deep learning algorithms and the technique used for identifying significant features. The comparison of model performances and optimization is elaborated in Section IV. Last but not least, Section V. discusses the final model framework, strategic maintenance management and describes novel contributions before a conclusion is made in Section VI.

TABLE 1. Medical device performance prediction studies.

Method/Equipment	Infant incubator	Defibrillator	Infusion or Perfusor pump
Author/ Year	Spahić, et al. [20] and Kovačević, et al. [21]/2020	Badnjević, et al. [22]/2019	Hrvat, et al. [19]/2020
Parameter	measurement error, number of parts, utilization coefficient, maintenance interval	device parameters, safety and performance parameters (output energy, age, manufacturer, type etc.) - 38 features	visual inspection, performance measurement (usage rate, environmental condition, age) inspection result
Data	three years period, 137 units at Bosnia and Herzegovina	three years period, 1221 units at Bosnia and Herzegovina	five years period, 1738 units at Bosnia and Herzegovina
Method	Feedforward Artificial Neural Network with Fuzzy Classifier	Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine and Naïve Bayes	Artificial Neural Network (feedforward) and Recurrent Neural Network (feedback)
Prediction	functional (accurate) or non-functional (faulty)	pass inspection (positive) or fail inspection (negative)	pass or fail
Accuracy	100% accuracy	100% accuracy	98.41% accuracy
Other performance metrics	100% specificity, 100% sensitivity	100% precision	not specified

II. RELATED WORKS

Machine learning (ML) studies statistical models that create mathematical models using sample data to make predictions and are widely used, especially during post-pandemic [14]. These models are based on probability, statistics, and algorithms. Contrarily, deep learning (DL) is a subset of machine learning which uses neural network topologies with an input layer, an output layer, and several hidden layers as a whole system. Like ML, DL can be classified as supervised and unsupervised learning. Examples of supervised learning in DL are Artificial, Convolutional, and Recurrent Neural Networks [15].

Researchers highlight the medical device reliability topic into three main areas, namely, i) risk management, ii) performance prediction for medical devices using ML, and iii) medical device management systems. Risk management areas can be classified into risk analysis, failure, and prioritization of medical devices' reliability. A combination of Failure Mode and Effect Analysis with Fuzzy (FFMEA) technique [16], FMEA [17], and Analytical Hierarchy Process (AHP) [18] are among the methods discussed in the risk management area.

Table 1 tabulates the utilization of machine learning in predicting medical device failure. The evaluation of the performance of medical devices is discussed; however, these studies only apply to one type of equipment. Three pieces of types of equipment involved were an infusion and perfusion pump by Hrvat, et al. [19], an infant incubator by Spahić, et al. [20], Kovačević, et al. [21] and a defibrillator by Badnjević, et al. [22].

In the mentioned studies, Kovačević, et al. [21] predicted device functionality for infant incubators in two categories: accurate and faulty, with a 98.5% accuracy and for decision trees algorithm and a 100% accuracy for artificial neural network algorithm [20]. Badnjevic et al. investigated mechanical ventilators and infant incubators using a similar approach. A defibrillator's performance parameter value was used to predict whether equipment would pass (positive) and faulty (negative) inspection using a random forest classifier with good accuracy of 100% [22].

On the other hand, Hrvat, et al. [19] achieved 98.41% accuracy using conformity evaluation, where the results are classified as pass or fail for infusion and syringe pumps. These findings suggest that even if a model for performance prediction has been developed, the model does not apply to other categories of critical medical equipment. Furthermore, there is a lack of an approach using cost analysis that would impact the existing maintenance programme.

Medical devices management system is an administration point of view topic to ensure the device's dependability. The factors affecting the maintenance and management of medical equipment were identified by Bahreini et al. [23] and are summarized in the following categories: human resources, financial, resources, inspection and preventive maintenance, physical documentation, education, service, quality control, information bank, management, services, training and education, design and implementation.

Moreover, prioritization is an alternative to overcome limitations in maintenance tasks and replacement plans. Maintenance and replacement expenses are optimized by categorizing medical devices into several criticality categories [18]. Recently, Zamzam et al. [24] proposed three robust models: corrective, preventive, and replacement plans for medical equipment at health clinics. The equipment is classified into low, medium, and high categories using the k-means technique. The result concludes that Support Vector Machine (SVM) outperforms other algorithms in prioritizing medical devices, with an accuracy of 99.42% for the preventive maintenance model.

There is still an opportunity for improvement where the model can be expanded to include critical medical devices at larger healthcare institutions that provide more comprehensive clinical services. In addition, the study excludes high-end equipment with high maintenance costs because such equipment is only available in larger facilities, such as hospitals. The study may be improved by using unstructured data to uncover other critical aspects of failure prevention.

The SVM approach in ML was also chosen in other medical device studies to assess the quantity of repairs and time-lapse to subsequent repair or failure [25]. Nevertheless, the authors developed the model based on a single SVM approach and limited assessment of infusion pumps and pulse oximeter device maintenance records.

Meanwhile, from a reliability engineering point of view, DL is used for optimal maintenance strategy in estimating Remaining Useful Life (RUL) for bearings using sensors [26]. A heterogeneous sensor framework is suggested for maintenance decisions with RUL prognostics estimation and post-prognostics verdict [27]. Besides, a DL technique is used to build an intelligent fault diagnosis using a gear-box dataset and bearing dataset to validate the performance of the machines [28], and a time-series application is applied as a case study for a cylinder of a small trolley in the automobile assembly line. The results benefit the manufacturing industry in improving the existing practices [29].

None of the published works utilizes DL to predict the need for medical device maintenance. In addition, existing studies utilize only structured numerical data, and other researchers have not investigated the application of multimodal or narrative text data. The unstructured text data contains valuable information regarding the primary cause of equipment failures.

As a result, to address the shortcomings in the aforementioned studies, this research will investigate the following new scientific findings:

- The existing published works on medical equipment performance prediction are for individual devices: defibrillators, infusion pumps, and infant incubators [19], [20], [22], [25]. The prediction of performance includes faulty and accurate or pass-and-fail responses. However, comprehensive studies are still lacking, and cost impact is not described. This study offers comprehensive cost analysis and the development of a new predictive model; (i) class 1, unlikely to fail in the first three years, (ii) class 2, likely to fail in three years; and (iii) class 3, likely to fail after three years from its commissioning date.
- The related works analysis concludes that medical device reliability can be classified into three distinct areas: management system, performance prediction, and risk assessment. In most studies, failure code analysis was proposed for reliability assessments. However, the studies still require manual human intervention and expert personnel to perform manual and thorough inspections.
- In addition, none of the published works demonstrates the usage of text narrative input for medical device performance prediction. Thus, this research proposes a new approach to medical device failure prediction by leveraging multimodal data. The proposed approach integrates structured data and text narratives (unstructured data). Besides, the word cloud is used to determine the root cause of failures to strengthen the existing routine maintenance duties.
- In accordance with the standard procedure, the cost of scheduled and corrective maintenance for medical equipment will be incurred following their acquisition, regardless of when the failure may occur. Government healthcare facilities invest significantly in high-tech medical equipment, necessitating a contract obligation for routine and corrective maintenance. By incorporating the prognostic aspect of AI, it is possible to anticipate the failure of a device and prepare the necessary resources.
- Besides, the proposed approach will enable data-driven decision-making. By analyzing vast datasets, AI-enabled predictive maintenance systems generate valuable insight. These insights can be used to optimize maintenance processes, refine purchasing decisions and inform strategies or device replacement.

Therefore, by utilizing data-driven decision-making using AI, healthcare organizations can make more informed decisions, streamline operations, and improve long-term planning. The ML and DL algorithms are tested to develop the best multimodal predictive model. The framework will be deliberated in the next section.

III. METHODOLOGY

This section comprehensively describes the proposed predictive models' development and associated techniques. All algorithms are explained, and details on the entire process of the proposed framework are included.

A. DATASET DESCRIPTION

In this study, we constructed a model of medical equipment failure based on data acquired from five hospital classes under Malaysia's Ministry of Health hospitals in Perak. The selected hospitals are state, major, minor, non-specialist, and special psychiatric facilities. Perak is a state located on the Malay Peninsula's western coast. It shares land borders with the Malaysian states of Kedah, Penang, Kelantan, Pahang, and Selangor to the north, northwest, east, and south, respectively.

Perak is one of the largest states in Malaysia. As of June 2022, the population is 2,500,000.00, with a land area of 20,099 km^2) [30]. 43.50% of the total 8,294 critical medical devices included in the study are located in the Kinta area, followed by 18.57% in the Larut Matang district and 11.4% in the Hilir Perak district.

The Perak State hospital, as tabulated in Table 2 has 1,394 beds, two major specialist hospitals with 608 and 548 beds, two minor specialists with 305 and 250 beds, nine non-specialist or district hospitals with 50 to 160 beds, and an institution for specialized psychiatry with 1800 beds. A total of 15 hospitals involved with clinical services are

TABLE 2. Selected healthcare facilities.

No.	Type and Cluster Division	Number of Beds	Number of Equipment
1.	State Hospital (Center)	1,394	3,236
2.	Major Specialist (Northern)	608	1,414
3.	Major Specialist (Northwest)	548	951
4.	Minor Specialist (Southern)	250	597
5.	Minor Specialist (Southwest)	305	667
6.	Non-specialist (Center 1)	160	204
7.	Non-specialist (Centre 2)	93	126
8.	Non-specialist (Northern 1)	136	165
9.	Non-specialist (Northern 2)	100	136
10.	Non-specialist (Northern 3)	90	126
11.	Non-specialist (Northern 4)	75	165
12.	Non-specialist (Southern 1)	100	120
13.	Non-specialist (Southern 2)	68	119
14.	Non-specialist (Southwest)	50	100
15.	Psychiatric Institution (Center)	1,800	168
	Total	5,777	8,294

managed by clustering in their respective areas, and the state hospital delivers 15 speciality services and designated sub-specialists based on the region. General Medicine, General Surgery, Pediatrics, Orthopedics, Obstetrics, Ophthalmology, ENT (Otorhinolaryngology), Emergency Medicine, Psychiatry, Dental, Dermatology, and Nephrology.

Gigantic maintenance and device details are challenging to manage since they are too complex and collected by the current database management systems, namely Asset and Services Information System (ASIS). ASIS is a data source for the study and analysis, with data available from 1991 to April 2021 (30 years). Even though the system can hold a vast amount of data, challenges arose when clinical engineers could not convert a large amount of maintenance data into a reliable instrument where more strategic maintenance plans could be implemented.

The Malaysia Standard (MS2058:2018) is the standard for active medical devices deployed in healthcare facilities, and maintenance tasks are described to prolong the device's useful life. In the standards mentioned above, laboratory, diagnostic and therapeutic are the three main categories that cover the industry's needs. These categories are grouped into patient support devices and critical devices. This study focuses on critical medical devices to accommodate critical needs; research on non-critical devices is accessible in current knowledge.

Therefore, medical equipment under the laboratory category is excluded and is classified as non-critical devices. A total of 8,294 pieces of equipment with 44 types are applied: 36.23% are infusion pumps, and 15.18% are physiologic monitoring systems for acute care. Other devices are below 10% in percentage, comprehensive of high-end equipment. Most high-end equipment includes Magnetic Resonance Imaging, Computed Tomography, Radiographic or Fluoroscopic Systems, Mammographic and Angiographic Systems.

Besides, other critical equipment such as ultrasonic, stimulators, resuscitators, peritoneal dialysis units, pacemakers, lithotripters, injectors, surgical hand drills, dental radiographic, mobile radiographic/fluoroscopic, laparoscopes, colposcopes, colonoscopes, cystoscopes, and drills bone are also included in the dataset.

B. STRUCTURED DATA PRE-PROCESSING

The data collection is extracted from ASIS, and pre-processing is performed to select significant input parameters, normalize the data, and exclude the missing data. Normalization ensures all characteristics are of the same scale and range. The dataset's vector-wise $z - score$ is returned after normalization, with a standard deviation of one and a center value of zero [24].

The $z - score$ is determined by measuring the separation of each data point from the mean and standard deviation. The $z - score$ value is represented by x, μ is the mean, σ is the standard deviation, and n denotes the highest probability estimation of the population's standard deviation. The standard deviation of x and the mean of x, will be returned as a vector and matrix, respectively. The following equation is used to calculate the $z-score$ of an x value after normalization, which operates separately on each column of data using equation (1) and (2) below:

$$z - score = \frac{x - \mu}{\sigma} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_1 - \mu)^2}{n}} \tag{2}$$

C. UNSTRUCTURED DATA PRE-PROCESSING

The pre-processing process is conducted on text narrative data to eliminate insignificant words such as a personnel name, 'and', 'for', 'ok', 'to', 'done', 'good', 'functioning' etc. Other pre-processing tasks are to change to lowercase since the capitalization has often been carried out inconsistently and remove leading or trailing white space. The punctuation and non-letter are removed. Any infrequent words that occur less than two times are eliminated to simplify the word cloud. The Bag of Words (BOW) technique generates the word cloud by cluster using the imported text narrative information.

The process involved treating n-grams as a specific word, and the model was fitted using a bag-of-n-grams model. The tokenized breaks the words into tokens, and BOW counts the unique words. The word cloud creation tools include the LDA for text document analysis [13], [31], as in Fig.1. The model supposes the topic mixtures, and these topics are based on Dirichlet distribution with concentration parameters of α correspondingly. Besides, z represent the random variable depicted by an integer from one through K, and w will characterize an integer from number one through the sum of words in the entire vocabulary [32].

Thus, the LDA fusion and a perplexity modelling is generated, where θ_1 to θ_D is the pool of D documents which is

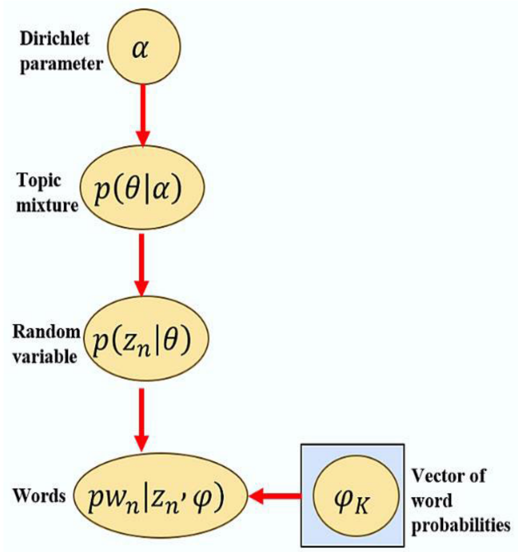

FIGURE 1. Latent Dirichlet Allocation Fusion.

known as a topic mixture, and over K, number of topics. It is depicted by the vectors of word probabilities φ_1 to φ_K. using equation (3) below:

$$p\left(w\mid\alpha,\varphi\right)=\int_{\theta}p(\theta|\alpha)\prod_{n=1}^{N}\sum_{z_n}p(z_n|\theta)pw_n|z_n,\varphi)d\theta \quad (3)$$

D. PREDICTIVE MODELLING

The response boundary classes for Classes 1, 2 and 3 are identified. The class is divided using an arbitrary technique with a balance of numbers between classes based on the pattern in the data. The features that ML and DL algorithms apply are deliberated in this section.

1) FEATURES

After the LDA cluster is determined, the cluster data is converted to numerical data and merged with structured data. ML and DL techniques are applied to the dataset, and the results are compared.

A total of 18 features are used to predict the time to the first failure. There are 17 features of structured data and one feature of unstructured data adopted from maintenance notes. Of the total 17 features, six categorical features are tabulated in Table 3 consisting of various hospital codes, equipment types, country of origin, manufacturers, brands and more than a thousand models in the dataset.

Meanwhile, another eleven features are converted to numerical and are normalized to ensure the data are on the same scale. The nature of each feature is explained in the description column in Table 4 associated with their respective values.

2) MACHINE LEARNING ALGORITHMS

Four supervised machine learning was utilized in this classification problem. The ML algorithms are SVM, Decision Trees (DT), Naïve Bayes (NB) and Ensemble Classifiers (EC). First, the SVM algorithm seeks a separating hyperplane between two classes with the best performance. The hyperplane with the most significant margin between the two classes is the optimum hyperplane for an SVM.

Decision Trees are the next method used in creating the model, as they are easier to interpret, have faster training, and are less complex. This algorithm's tweaking can be accomplished by adjusting the maximum number of splits. The branch is extended to the leaf node of the tree, where the response resides.

Gaussian and Kernel Naive Bayes are two Naive Bayes algorithm variants. The Naïve Bayes algorithm is based on Bayes Theorem, and its applicability is contingent upon the training data. The model is trained on four ML algorithms above using the five k-fold cross-validation technique with 80% training and 20% testing segregation with random fold separation, as in Fig.2.

TABLE 3. Categorical features.

No.	Predictor	Values
1.	Hospital code	15 different codes (E.g., PRK300, PRK301)
2.	Type description	44 types (E.g., Aspirators)
3.	Country	37 countries (E.g., Malaysia)
4.	Manufacturer	511 manufacturers (E.g., Datex Ohmeda)
5.	Brand	568 brands (E.g., Smartvent)
6.	Model	1,375 models ((E.g., Smartvent 7900)

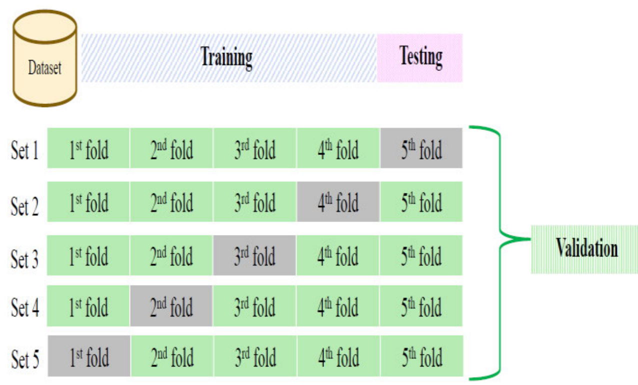

FIGURE 2. K-Fold Cross Validation for ML.

3) DEEP LEARNING ALGORITHMS

A neural network algorithm is more complex and has harder interpretability. Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are frequently employed in engineering applications. ANN algorithm is more complicated to interpret. The classifier options include narrow, medium, wide, bilayered, and trilayered and are based on a feedforward neural network.

TABLE 4. Numerical features and designated values.

No.	Predictor	Values	Description
1.	Age	Unique values	Date equipment is purchased until April 2021
2.	Service support	1: End of production, 0: Service available	Service availability after medical devices are purchased. A discontinuation letter is issued when parts are no longer available
3.	Asset condition	0: Active/in use, 1: Unrepairable failure but still in use, 2: Approved for disposal	Equipment condition and evaluation for disposal
4.	Service intention	1. Diagnostic, 2. Therapeutic,3. Life Support	Primary function
5.	Frequency maintenance requirement	1: PPM, yearly, 2: PPM, twice- yearly, 3: PPM with Quality Control Certificate	Maintenance interval and quality control measure the radiation effect on imaging equipment
6.	Maintenance complexity	1. Average maintenance with EST, 2. High-end maintenance with EST	EST measure the leakage current for equipment that is directly attached to patients
7.	Total downtime	Unique values	Count from initial failure until the device returns to its working state (in hours)
8.	Alternative and backup	0: No loaner provided, 1: Loaner provided	Temporary replacement during a breakdown
9.	Operations	1: 12 hours/6 days, 2: 24 hours/7 days	Average utilization based on location
10.	Maintenance cost	Unique values	Maintenance cost for parts and labour.
11.	Purchase date	Unique values	

The CNN layers execute the operations using an activation or rectified linear unit (ReLU) layer, convolution, and pooling layers. As for the output layers, the classification group comprises fully connected layers that produce a K-dimensional vector of predicted classes (three response classes), a softmax layer and a classification layer [33].

Meanwhile, RNN is unique due to their hidden state and loops. A Long Short-Term Memory Network (LSTM) is the most common type of RNN. In LSTM networks, extra gates regulate which data from the hidden cell is sent to the output and the subsequent hidden state. An additional to the LSTM layer is the number of hidden units.

4) DEEP LEARNING OPTIMIZERS

Deep learning networks are developed with layers and a dedicated optimizer. Three optimizers are involved: Root Mean Square Propagation (RMSProp), Stochastic Gradient Descent with Momentum (SGDM), and Adaptive Moment Estimation (Adam). SGDM optimizer can oscillate along the path of steepest descent, leading to the best result.

The SGDM update can be determined by equation (4), with γ as the current iteration's contribution from the previous gradient step, α representing the learning rate, ℓ as the iteration number, θ as a parameter vector, and $\nabla E(\theta)$ is the loss function. Besides, RMSProp maintains a moving average of the parameter gradients' element-wise squares using a mathematical computation with β_2 as the moving average's decay rate, and ϵ denotes the added small constant to prevent zero division as in equation (5).

Meanwhile, Adam optimizer uses a parameter update similar to RMSProp, but with an additional momentum term. It maintains a moving average of the parameter gradients and their squared values using element by element approach and gradient decay rate of β_1 as shown in equation (6) [34]:

a) SGDM:

$$\theta_{l+1} = \theta_\ell - \alpha \nabla E(\theta_\ell) + \gamma(\theta_\ell - \theta_\ell - 1) \tag{4}$$

TABLE 5. Five metrics for evaluation.

No.	Performance Indicator	Equation
1.	Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
2.	Sensitivity/Recall	$\frac{TP}{TP+FN}$
3.	Specificity	$\frac{TN}{TN+FP}$
4.	Precision	$\frac{TP}{TP+FP}$
5.	F1 Score	$2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall}$

b) RMSProp:

$$v_\ell = \beta_2 v_{l-1} + (1-\beta_2)\left[\nabla E\left(\theta_\ell\right)\right]^2 \tag{5}$$

c) Adam:

$$m_\ell = \beta_1 m_{l-1} + (1-\beta_1)\nabla E(\theta_\ell) \tag{6}$$

E. MODEL PERFORMANCE METRIC

The accuracy of the model's predictions depends on the standard of the input characteristics used by the algorithm. The evaluation is performed using the confusion matrix. Diagonal cells express correctly classified class, whereas off-diagonal cells express the opposite. The recall indicator gauges the reliability of the result or positive outcomes for a true positive rate, whereas precision measures positive predictive values with correct prediction [35]. The integration of recall and precision is examined in the F1 Score indicator, where this harmonic average indicates the model's dependability.

The metrics are calculated by adding the correctly classified values divided by the total number of observations, as illustrated in Table 5. TP represents a true positive, the true negative is TN, and false values are symbolized by the letter F [24]. These values are obtained through a confusion matrix with a 3 × 3 matrix where the x-axis is the predicted class, and the y-axis is the true class [36].

F. SENSITIVITY ANALYSIS

A Leave-One-Out, Minimum Redundancy Maximum Relevance (MRMR), Chi2, ANOVA, and Kruskal Wallis are the techniques used in the feature selection stage. Continuous and categorical features are supported in MATLAB and applied to all methods. First, the leave-one-out technique is used; the error is calculated and then divided with all feature errors to obtain the ranking ratio [37], [38], [39]. The leave-one-out technique is performed by removing one-by-one features and organized in descending order to highlight the impact of the features on the model.

The MRMR algorithm identifies the best possible set of maximally and reciprocally different features and may accurately describe the response variable. The technique increases the relevance of a feature set to the response variable while minimizing its redundancy; minimizes W_S and maximizes V_S in equations (7) and (8) where x is the selected feature, y is the response variable, S represent a subset of the desired feature, and |S| is the number of features in S [40]:

$$\max V_S = \frac{1}{|S|}\Sigma_{x\epsilon S} I(x, y) \tag{7}$$

$$\min W_S = \frac{1}{|S|^2}\Sigma_{x,z\epsilon S} I(x, z) \tag{8}$$

A Chi2 algorithm uses individual chi-square tests to determine whether each predictor variable is independent of the response variable. Then p-values of the chi-square test statistics are used to rank the features. A low p-value denotes the predictor variable is reliant on the response variable or is categorized as a significant feature. The score relates to -log(p), where p is the number of predictors. A predictor with a high score value is essential for this algorithm.

A similar approach is applied to the ANOVA algorithm where -log(p) shows the score through a one-way analysis of variance. The one-way analysis of variance is performed for each predictor variable, sorted by class, and then features are ranked using the p-values. In the meantime, the Kruskal-Wallis test is a nonparametric variation of the Wilcoxon rank sum test that extends to more than two groups. The p-value calculates the significance of the chi-square statistic, which takes the role of traditional one-way ANOVA [41]. All these feature selection techniques are compared, and the most significant features are determined in the result section.

IV. RESULTS

The results section portrays the overall outcomes of this research. All algorithms' integration and optimization stages are described. The sensitivity analysis approach minimizes the model complexity and shortens the training time.

A. PERPLEXITY MODEL

An extended bag of words or a topic model: LDA technique is applied for clustering and classifying the pattern on maintenance notes or unstructured data. This topic modelling discovers the hidden patterns and identifies the main themes or the semantic topics. The maintenance notes are notes written by technical personnel after attending a breakdown event.

A perplexity graph is plotted as in Fig.3, where the graph indicates that 15 main topics are sufficient to be analyzed before the graph decreases. The perplexity approach reveals how accurately a model explains the data. A better-fit model has a lower perplexity. The imprecise number of clusters selection will result in an inaccurate interpretation of the main topics.

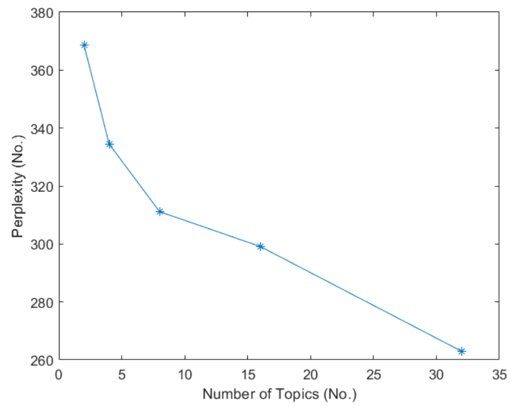

FIGURE 3. Number of Topics for Perplexity Model.

The 15 clusters illustrate the pattern in the unstructured data, and the word clouds are developed. The 15 cluster groups are summarized into six important labels by removing the duplicates, and the clusters are mapped back to the dataset. After clustering, the data is labelled into six groups: repair, replace, visual inspection, vendor, and beyond economic repair. The repetitive groups are simplified from all the clusters, and six top clusters are established as follows:

1) Zero failure: Maintenance notes not available or action taken is not specified.
2) Repair: Inclusive of calibration work and major troubleshooting
3) Replace: Maintenance work performed that involves changing parts or components in the device
4) Visual Inspection: Physical inspection or only minor troubleshooting is required during the assessment
5) Vendor: Unrepairable failures by the maintenance team are returned to the vendor. This cluster includes devices under warranty.
6) BER: The term used in the ASIS for devices exceeding their lifespan, reaching the end of life, or an unrepairable failure and proposed for disposal.

B. TIME TO FAILURE PREDICTION

Classes 1,2, and 3 are identified based on the time to the first failure event. Class 1 is defined as unlikely to fail within the first three years from the purchase date, while Class 2 is for devices that are likely to fail within three years. Class 3 is for devices likely to fail more than three years after purchase. This section compares four ML techniques in Table 6, ANN in Table 7, CNN in Table 8 and LSTM in Table 9 using all 18 features of multimodal data.

The result for ML in Table 6 concludes EC algorithm denotes the highest accuracy and recall of 87.90%, 88.21%, 93.95%, and 88.06% for precision, specificity, and F1 Score, respectively. The DT and SVM algorithms achieve 85.00% and 82.80% accuracy, and the performance is slightly lower than the EC algorithm.

TABLE 6. Performance evaluation for ML.

No.	Performance	DT	NB	SVM	EC
1.	Accuracy	85.00	69.70	82.80	87.90
2.	Recall	85.00	69.71	82.77	87.90
3.	Precision	85.05	73.08	85.17	88.21
4.	Specificity	92.50	84.85	91.39	93.95
5.	F1 Score	85.02	71.35	83.96	88.06
6.	Training time (sec)	34.97	4864	55732	1660.5

As for the ANN network in DL, the model performs the best with 81.80% accuracy, as shown in Table 7. However, the specificity performance for ANN reaches 90.90% compared to 93.95% for the EC model. Like ML algorithms, the ANN network requires significant training time to deploy the model.

TABLE 7. Performance metrics for ANN.

No.	Performance	ANN
1.	Accuracy	81.80
2.	Recall	81.80
3.	Precision	81.87
4.	Specificity	90.90
5.	F1 Score	81.84
6.	Training time (sec)	1.1771e+05

The CNN network performance is tabularized in Table 8, with the SGDM optimizer achieving the best performance with 80.77% accuracy compared to RMSProp and Adam optimizers. SGDM requires a moderate training time of 44 seconds, slightly higher than RMSProp. Adam optimizer demands the longest training time for the CNN network. The CNN network appears to have less complexity and less time consumed.

A small difference in performance is observed for the LSTM network in Table 9 compared to CNN. Both CNN and LSTM networks share the characteristic that SGDM achieves the highest levels of performance in comparison to RMSProp and Adam optimizers. The application of optimizers in DL suggests that CNN performs better than LSTM. Although both accuracies are closed, CNN performs better in a training time of 44 seconds compared to 135 seconds using LSTM. Additionally, DL requires significantly less time spent training, one of its primary advantages. DL finishes the training procedure in 44 seconds for CNN, while it takes ML 1660.5 seconds for EC to complete.

TABLE 8. Performance metrics for CNN.

No.	Performance	RMSProp	SGDM	Adam
1.	Accuracy	79.14	80.77	78.08
2.	Recall	75.48	76.95	77.90
3.	Precision	74.81	76.89	77.83
4.	Specificity	87.57	88.52	89.43
5.	F1 Score	75.14	76.92	77.86
6.	Training time (sec)	42	44	93

TABLE 9. Performance metrics for LSTM.

No.	Performance	RMSProp	SGDM	Adam
1.	Accuracy	78.57	80.60	79.46
2.	Recall	77.44	77.12	79.79
3.	Precision	77.18	77.45	79.73
4.	Specificity	88.84	88.49	89.76
5.	F1 Score	77.31	77.28	79.76
6.	Training time (sec)	59	135	56

C. MODEL OPTIMIZATION

The parameter tuning is performed on the model to introduce a unique identity and maximize accuracy. The tuning options are distinctive for every algorithm and are conducted using possible parameters until the optimal results are obtained.

Table 10 indicates the hyperparameter tuning conducted for four ML algorithms. For DT, adjusting the maximum number of splits increased the training time but improved the accuracy. Similarly, the kernel tuning option for NB and SVM increased the time to produce a higher accuracy. The decreased number of splits to 20 increased the training time

TABLE 10. Optimization for ML.

Algorithm	Optimization	Training time (sec)	Accuracy
DT	The maximum number of splits is adjusted between 100 and one	Increased from 4.42 to 34.97	Increased from 83.30% to 85.00%
NB	Optimized from kernel to optimizable tuning	Increased from 89.895 to 4864	Increased from 69.00% to 69.70%
SVM	The kernel scale is optimized from an automatic scale to one	Increased from 315.78 to 55732	Increased from 79.10% to 82.80%
EC	Bagged to Adaboost and changed the number of splits from 9,820 to 20, with 30 learners and 0.1 for the learning rate.	Increased from 52.58 to 1660.5	Increased from 87.70% to 87.90%

and accuracy for EC. ML optimization usually increases the training time and growth of the model complexity.

The optimization for ANN algorithm is demonstrated in Table 11. Optimizing the layer sizes reduces the model intricacy and training time. The hyperparameter tuning is conducted to all first, second and third layers before the best model is deployed. The finest accuracy for ANN is 81.80% after optimization.

TABLE 11. Optimization for ANN.

Algorithm	Optimization	Training time (sec)	Accuracy
ANN	First layer size from 100 to 10,10,10 for a first, second and third layer.	Reduced from 5201.8 sec to 1.1771e+05	Increased from 80.90% to 81.80%

The optimization for CNN models varies by application of mini-batch size (MBS) and epochs. Fig.4 exhibits the result for the CNN network using the RMSProp optimizer. The application of 300 MBS produces the greatest accuracy of 79.14%.

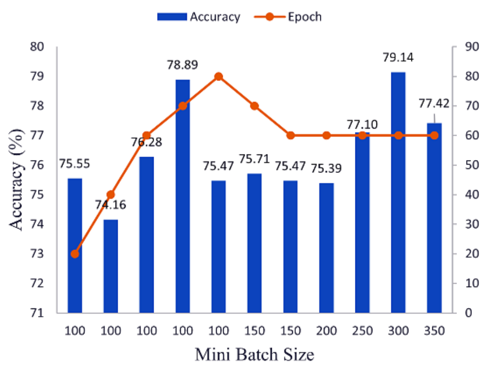

FIGURE 4. Optimization for CNN with RMSProp Optimizer.

As for SGDM and Adam optimizers, smaller sizes, which are 200 and 100, are required accordingly. The 200 MBS and epochs of 60 for SGDM in Fig.5 produce the highest accuracy of 80.77%.

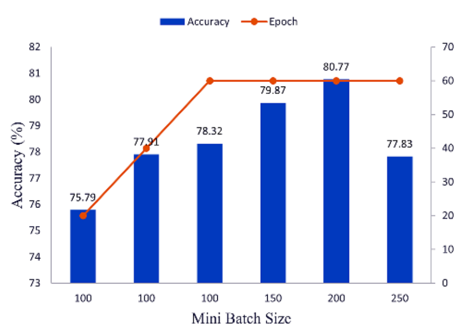

FIGURE 5. Optimization for CNN with SGDM Optimizer.

Besides, an Adam optimizer requires a lower number of MBS. The setting of 100 MBS with 20 and 60 epochs in Fig.6

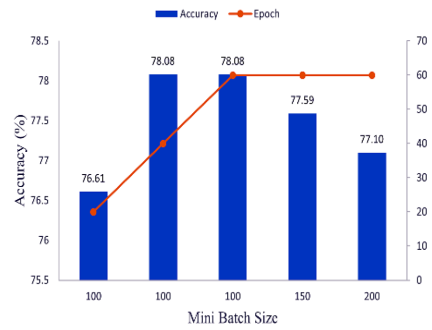

FIGURE 6. Optimization for CNN with Adam Optimizer.

produces the same accuracy. The Adam optimizer denotes an accuracy of 78.08%, the lowest compared to SGDM and RMSProp optimizers. The result concludes the optimal values for CNN networks are 60 epochs with RMSProp (MBS 300), SGDM (MBS 200), and Adam (MBS 100).

In the meantime, the LSTM setting parameters are tuned by changing the number of hidden units in the network. Fig.7 shows the impact on the accuracy when the number of hidden units varies for the SGDM optimizer. The increased value of 200 to 300 for hidden units in the network raises the accuracy from 76.53% to 80.60%. However, more than 300 hidden units will continue to reduce the performance in accuracy.

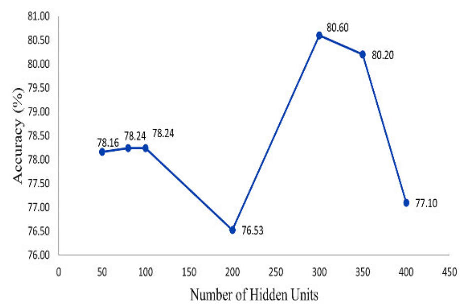

FIGURE 7. Number of Hidden Unit Optimizer in LSTM.

Comparing all three optimizers in LSTM, Fig.8 describes the impact of varying MBS on the networks. During optimization, various MBS slightly differ in training and validation accuracy. The highest accuracy is found when the epoch is set to 60. By applying 60 epochs in the networks, the SGDM optimizer has the highest accuracy of 80.60% when the MBS is set to the value of 100. For Adam and RMSProp, MBS of 400 gives the best result in accuracy, equivalent to 79.46% and 78.57%, respectively. The result concludes an optimal value for LSTM networks is 60 epochs, 300 hidden units with SGDM (MBS 100); Adam and RMSProp (MBS 400) to produce the best result.

D. SENSITIVITY ANALYSIS RESULT

The Ensemble Classifier model achieves the best accuracy of 87.90% after optimization compared to other DL techniques. Thus, the sensitivity analysis for ML is conducted

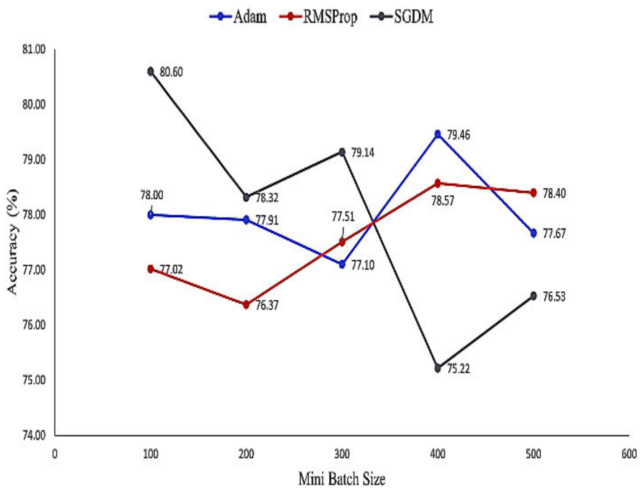

FIGURE 8. Mini Batch Size for LSTM.

using five techniques. Every technique requires a different analysis before the final significant features can be listed. The techniques applied to the ML algorithm are Leave-one-out, MRMR, Chi2, ANOVA, and Kruskal Wallis.

The leave-one-out method by a ranking ratio is illustrated in Table 12. The analysis requires eliminating one by one features, and the ratios are determined by calculating the error obtained from the confusion matrix. The 18 features are excluded individually, and the ratios are calculated to determine the significant features.

TABLE 12. Sensitivity analysis using leave-one-out for ML.

Rank	Detract	Ratios
1.	Total downtime	1.391
2.	Maintenance notes	1.224
3.	Country	1.070
4.	Type description	1.055
5.	Maintenance complexity	1.033
6.	Operations	1.033
7.	Hospital code	1.030
8.	Maintenance cost	1.018
9.	Service intention	1.018
10.	Manufacturer	1.016
11.	Age	1.012
12.	Purchase date	1.010
13.	Brand	1.001
14.	Service support	1.000
15.	Alternative and backup	0.996
16.	Asset condition	0.983
17.	Frequency maintenance requirement	0.982
18.	Model	0.957

Out of the 18 features in ML, four features are not considered with less than 1.000 values: alternative backup, asset condition, frequency maintenance requirement and model. After eliminating the four insignificant features, the analysis is repeated for the remaining 14 features. All 14 features are selected having a ratio greater than 1.000. As a result, the accuracy of ML algorithm is increased from 87.90% to 88.30%.

In addition, for MRMR technique in Table 13, eight features are identified as the most important since the increased number of inputs will reduce the accuracy. Similarly, for Chi2 and Kruskal Wallis, six features are listed as substantial before the accuracy drops, with the highest accuracy of 87.80%. Meanwhile, ANOVA technique has eight most significant features with 88.00% accuracy.

Among all techniques, MRMR outperforms other methods with an accuracy of 88.80% compared to 87.90% before the analysis. The best MRMR model listed the eight most significant features: total downtime, maintenance complexity, maintenance notes, maintenance cost, age, purchase date, type description, and country.

The MRMR algorithm denotes the highest performance by identifying the best possible set of maximally and reciprocally different features to portray the response. The ability of the MRMR to minimize redundancy among features decreased eighteen to eight features, shortening the training time from 1660.5 to 901.66 seconds.

A predictive model using an Ensemble Classifier algorithm is proposed, and MRMR is the best technique for feature selection. The proposed model using MRMR increased performance: slightly higher than when all features were applied. The performance after the MRMR technique improved its percentage to 94.41%, 88.82%, 88.46%, and 88.84% for specificity, recall, precision and F1 Score as tabulated in Table 14. The final model optimization involves 20 maximum splits, 30 learners and a 0.1 learning rate to reach 901.66 seconds.

V. DISCUSSION

This section discusses the characteristics of the best classifier, including the parameter settings and receiver operating characteristics curves. A long-term benefit of the proposed model is discussed in this section, with the associated cost impact for the proposed model.

A. BEST CLASSIFIER CHARACTERISTICS

Receiver operating characteristics (ROC) curves show the performance of the multi-class classification model. A ROC curve compares the true positive rate (TPR, or sensitivity) to the false positive rate (FPR, or 1-specificity) for various classification score thresholds.

The one-versus-all coding design is used to identify the ROC curve for each class in this multi-class classification task. The technique views a multi-class classification problem as a collection of binary classification problems, wherein each binary problem assumes one class is positive and others are negative.

Therefore, the area under curve (AUC) measured the classifier performance where a good model has AUC close to 1. Three different ROC curves are plotted based on classes, as in Fig.9. The plotted curve is for the best Ensemble Classifier model with 88.80% accuracy and 94.41% for specificity. The shaded area around the three ROC curves demonstrates the 95% confidence interval for the model.

TABLE 13. Features ranking results using four techniques.

Rank	MRMR	Chi2	ANOVA	Kruskal Wallis
1.	Downtime	Age	Maintenance notes	Maintenance notes
2.	Maintenance complexity	Downtime	Purchase date	Purchase date
3.	Maintenance notes	Maintenance cost	Age	Downtime
4.	Maintenance cost	Purchase date	Downtime	Age
5.	Age	Maintenance notes	Service support	Maintenance cost
6.	Purchase date	Model	Country	Service support
7.	Type description	Other features are excluded, if added, will reduce the accuracy	Asset condition	Other features are excluded, if added, will reduce the accuracy
8.	Country		Type description	
9-18	Other features are excluded, if added, will reduce the accuracy		Other features are excluded, if added, will reduce the accuracy	

TABLE 14. Comparison between EC and MRMR tuning.

No.	Metrics	EC (All features)	MRMR (Important Features)
1.	Accuracy	87.90%	88.80 %
2.	Specificity	93.95%	94.41 %
3.	Recall	87.90%	88.82 %
4.	Precision	88.21%	88.46 %
5.	F1 score	88.06%	88.84 %
6.	Prediction speed (obs/sec)	~2300	~2100
7.	AUC	96.00%	93.00%
8.	Training time (sec)	1660.5	901.66

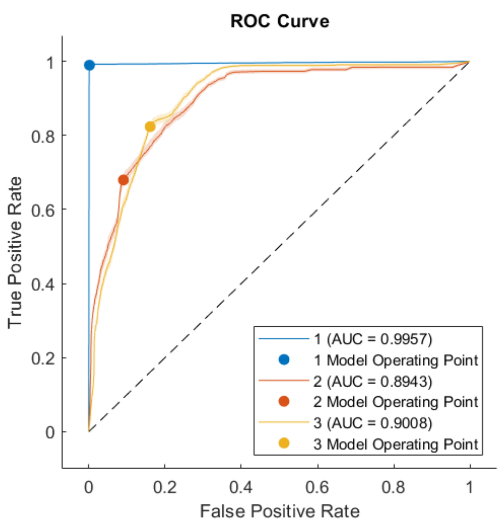

FIGURE 9. ROC Curves for the Best Classifier Model.

This confidence interval is obtained by bootstrapping the data using a specific function. The bootstrapping function is specified as a nonnegative scalar, and it uses the number of bootstrap samples to calculate the pointwise confidence intervals. The graph denotes a 95% certainty that the true parameter value lies within this interval. It can be seen that Class 1 has the best area under curve values of 99.57%, followed by 90.08% and 89.43% for Class 3 and Class 2.

TABLE 15. Classes description.

Classes	Class 1	Class 2	Class 3
Description	Device unlikely to fail within the first three years of purchase	Device failure within three years from the purchase date	Device failure after three years from the purchase date
Criticality	Low	High	Medium
Characteristics	Zero failure and zero parts cost	High number of failures	Moderate number of failures

B. STRATEGIC MAINTENANCE MANAGEMENT

An arbitrary technique is used to distinguish the devices into three classes. Classes 1, 2, and 3 are segregated based on the time to the first failure event. Class 1 is defined as unlikely to fail within the first three years from the purchase date, while Class 2 is for devices that are likely to fail within three years. Class 3 is for devices likely to fail more than three years after purchase. A strategic replacement plan should be in place to ensure efficient cost management and optimal budget utilization.

In the current practice, all 44 types of critical medical devices are included in a long comprehensive contract, and the monthly fees depend on the device's fee rate. The annual fee is calculated by multiplying the purchase cost by the mentioned fee rate. The fee rate is based on device criticality, which is 4.95%, 6.60%, 7.15%, 11.55%, 13.75%, and 19.25%. A higher rate is allocated for imaging devices due to its high maintenance cost and complexity. For improving and utilizing ML with a new comprehensive strategic maintenance management, the fee rate shall depend on the criticality of the classes.

Table 15 tabulates the class criticality into low, medium and high, representing time to the first failure. Class 2 should have a higher maintenance rate because it requires more rigorous maintenance than other classes and indicates the highest number of failures. Moreover, the services contract fee will not begin until the 4th year, and a substantial cost impact will be realized.

TABLE 16. Planned preventive maintenance plan for classes 1, 2 and 3.

No.	Criteria / Classes	Class 1	Class 2	Class 3
1.	PPM schedule	Revise from bi-annually to annually	Remain as suggested by the manufacturer	Revise from bi-annually to annually
2.	Maintenance task	Minimal	Stringent	Moderate
3.	Yearly maintenance budget allocation	Less	Priority	Moderate
4.	Loaner and rental budget allocation	Not required	Priority	Moderate
5.	Device warranty	Remains two years as normally offered by the manufacturer	Three years warranty after purchased (Inclusive of PPM and corrective maintenance)	Three years warranty after purchased (PPM only)
6.	Service contract	Starts at 4th year or later	Starts at 4th year	Starts at 4th year
7.	Cost impact	No additional cost for an extended warranty	Additional cost for extended warranty	Zero corrective maintenance or downtime cost in the first three years.
		Zero service contract cost for the first three years	Zero service contract cost for the first three years	Zero service contract cost for first three years; PPM is covered under warranty)

The criteria are separated into three classes in Table 16. This approach is proposed in lieu of the present practise of paying for scheduled maintenance after purchasing devices. The existing data indicate that Class 3 will fail after three years; hence a fourth-year service contract is recommended. As for Class 2, an extended warranty should be suggested, along with Planned Preventive Maintenance (PPM) and Corrective Maintenance (CM), and a service contract should be scheduled to begin in the fourth year.

The replacement plan is also strategized considering the number of ages and faulty equipment. With a high rate of failures throughout their lifespan, 57.19% of Class 2 equipment has been used for over ten years, as tabulated in Table 17. Replacement with new units should be considered to reduce maintenance costs due to ageing factors and significant failure events.

In a situation similar to Class 2, Class 3 has 65.18% of the equipment used for more than ten years and should be replaced after Class 2. Therefore, a comprehensive replacement plan policy shall consider both criteria, such as maintenance cost by service contract reaching the purchase cost and the priority for Class 2 replacement.

Furthermore, the Class 2 characteristic of first failure within three years after purchase, Class 2 has 173 devices over 20 years with 11-30 failures events, compared to 151 devices in Class 3. Similarly, 115 devices with more than 31 failures are observed for Class 2, compared to 29 devices in Class 3 equipment. The study proves that Class 2 has the highest priority for replacement based on age and number of failures compared to Classes 1 and 3.

Based on the overall analysis, Fig.10 demonstrates the sub-framework for Ensemble Classifier, which is the best predictive model after comparing ML and DL.

TABLE 17. Failure events based on classes.

Class	Failure event	Device's Quantities ≥ 20 years	≥ 10-19 years	≤ 9 years
1	None	32	157	2,042
2	1-10 events	126	716	1,160
	11-30 events	173	698	495
	≥ 31 failures	115	521	103
3	1-10 events	292	658	644
	11-30 events	151	131	35
	≥ 31 failures	29	14	2
	Total	918	2,895	4,481

The sensitivity analysis technique listed eight significant features for Ensemble Classifier with MRMR attained the finest after all approaches are compared. The sensitivity analysis reduced the model complexity and training time by minimizing the redundancy between features to achieve the objective.

Moreover, Fig.11 recaps the main outline of the model framework using multimodal from raw dataset to the integration of structured and unstructured data. The model optimization is conducted, and the model is evaluated before a final model is proposed.

C. COST IMPACT

Based on the literature findings, none of these nations uses ML to forecast the annual maintenance budgets. For example, life cycle cost estimation is applied in Saudi Arabia. A life cycle cost analysis in decision-making is based on total expenses rather than a device's initial purchase price [42]. In Turkey, they categorize medical devices into

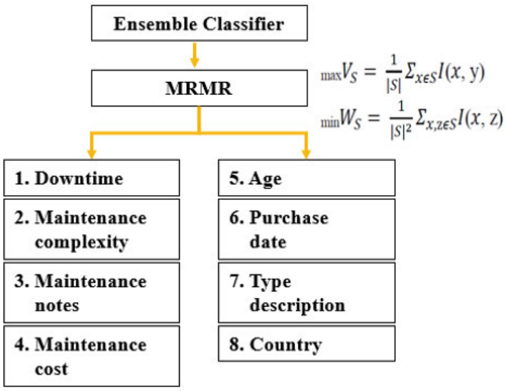

FIGURE 10. The Ensemble Classifier Model with MRMR Significant Features.

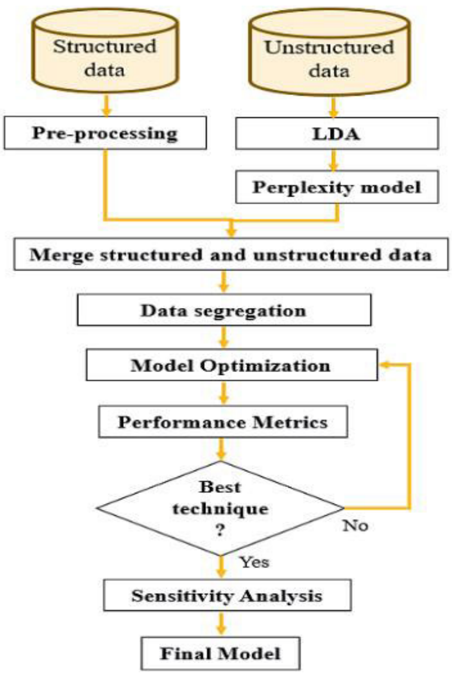

FIGURE 11. The Overall Proposed Framework of Multimodal Ensemble Classifier Model for Failure Prediction.

technological development groups such as low, medium, high and simple technology groups to determine the ratio for cost expenditure [43].

The replacement budget per year in the UK country is calculated based on total stock divided by lifetime. Fixed and operating costs are used in Nepal for financial management. Outsourcing is reported as the most cost-effective approach compared to in-house maintenance. For instance, the facility will anticipate average annual replacement expenses of one-tenth of the purchase cost if the equipment has a lifespan of ten years [44].

In New Delhi, cost expenditure is measured through the maintenance cost index, obtained from maintenance cost divided by capital cost [45]. The United States uses the cost-of-service ratio by dividing the annual maintenance cost total by the starting cost values. It serves as guidance for performance improvement, and the cost-of-service ratio in the United States ranges from 5% to 10%.

TABLE 18. Cost impact.

Class	Frequency	Current cost/year (MYR)	Long-term cost saving/year (MYR)	Saving (MYR)
1	PPM frequency changed from bi-annually to annually	2,838,246.45	2,638,990.00	199,256.45
2	PPM remains as existing			
3	PPM frequency changed from bi-annually to annually	2,440,926.61	2,313,852.18	127,074.43
			Total	326,330.88

Meanwhile, at the national level of Ethiopia in Africa, only 49.2% of hospitals had reported having a dedicated budget for purchasing new equipment [46]. Consequently, based on the findings, ML, DL and technological advancement are not currently utilized in maintenance and expenditure planning.

Implementing manufacturer suggestions on maintenance frequencies and procedures will not guarantee minimal failure. PPM requires USD 300 million per year to accommodate the needs of US hospitals. Yet, it is not promising uptime is guaranteed even though the PPM intervals are according to manufacturer recommendations [47].

The Joint Commission investigated the required practice maintenance instead of executing manufacturers' recommendations on procedures and frequencies [48]. The study concludes that even though the users differ in the procedures and maintenance frequencies, there is no evidence the method could negatively impact users and patients in American hospitals. Adaptability in scheduling and planning maintenance tasks is crucial when using predictive maintenance on medical equipment.

Due to the usage of patients and external control factors, planned maintenance tasks are frequently challenging to carry out at an appropriate time. It is a common event when the users are unable to release the equipment when it is in use or misplaced at another location. It impacted the PPM frequency and differed from the initial schedule. Reducing the frequency with regards to the necessity may accommodate the user and equipment needs for maintenance.

It is suggested to distinguish the budget for PPM and CM [49]. More accurate financial tracking between PPM and CM shall be established by segregating this. The price of a service contract can be negotiated with external service providers. These services can be obtained on a time and material basis or by entering a fixed-fee contract for a predetermined amount of time. The expense must be anticipated and included in relevant budgets in both scenarios.

TABLE 19. Root cause of failures for therapeutic equipment.

No.	Equipment	Failures components
1.	Aspirator	Suction, pump, tubing, and battery faulty
2.	Defibrillator	Battery, cable, lead, paddle, cuff, and monitor faulty
3.	Hand drill, surgical	Cable, battery, hose, leaking, and driver faulty
4	Hemodialysis unit	Diasafe filter, calibration needed, motor, alarm, leaking, pump, conductivity, tubing, valve, connector, diacap, and board faulty
5.	Incubator	Skin probe, probe, battery, temperature, power supply, screw, door, wheel, and fan faulty
6.	Infusion pump	Battery, sensor, power supply, board, cable, keypad, calibration needed, and loose screw
7.	Phototherapy	Bulb, photo light, cable, fused, and cable connection
8.	Ventilator	Oxygen and flow sensor, oxygen cell, battery, valve, calibration, cable, tubing, hose, leaking, filter, pressure, and board

The human resources cost can be reduced by minimizing the PPM schedule from bi-annually to annually for Classes 1 and 3. A specific percentage is set up for PPM and CM, with a higher allocation of 80% for corrective maintenance and another 20% for PPM [50]. Previous study concludes PPM demands 10% to 30% more of the overall budget than CM. A similar concept is proposed here where if the PPM is scheduled once yearly, a 20% budget allocation is proposed for PPM, and another 80% is allocated for CM.

Meanwhile, if the PPM is scheduled bi-annually, the portion is segregated to 70% CM and 30% PPM instead. By making this percentage as said, it is expected to achieve significant cost savings due to reduced resources required when the PPM frequency is reduced.

For Class 1, MYR199,256.45 is calculated as a cost-saving if the PPM frequency is reduced, as illustrated in Table 18. The fee rate for Class 1 shall be reduced due to its low complexity, and PPM can be replaced with a routine inspection where necessary. A service contract cost shall start from the 4^{th} year or later for Class 1 as the devices have recorded zero failures. Additionally, this equipment category requires a minimal budget for a loaner and rental fees during failure.

The most stringent class and has a higher severity is Class 2. The PPM schedule is suggested to remain as the manufacturer indicates since these devices are considered critical and can fail anytime after purchase. Since the failure of the purchased devices is expected to occur at any time, backup plans or the use of rented equipment must be arranged for an improved service.

Therefore, a higher budget should be allocated to Class 2 than Class 1 and 3. Besides, an extended three-year warranty is suggested to accommodate the needs of PPM and CM for Class 2. An extended warranty will eliminate the service contract cost in the first three years.

Devices under the Class 3 group are expected to fail after three years from their commissioning date. Hence, the approximate cost saving of MYR127,074.43 per year is expected to be achieved if the PPM schedule is reduced from bi-annually to annually. A service contract obligation shall be eliminated for the first three years for Class 3 since the failure is expected to occur after three years.

Consequently, an extended warranty of three years inclusive of PPM is adequate to cater to the needs of this group. Despite recommendations to reduce PPM frequency for Class 1 and 3, technical personnel shall intensify PPM duties to meet demands before the following PPM schedule and guarantee useful life of the devices is not compromised. According to IEC60601 and related standards, the functionality and safety of medical devices should be ensured during PPM.

This study predicts Classes 1, 2, and 3 represent time to the first failure. By categorizing devices into these classes using machine learning in decision-making, a new maintenance strategy is proposed using multimodal data for intelligent healthcare engineering. In the existing literature, Kovačević, et al. [21] explores the study of infant incubator devices where a different response or output is highlighted, which are accurate and faulty classes with a model obtain an accuracy of 98.5%.

A different device is utilized using a Random Forest classifier with the ultimate objective of predicting positive and negative responses representing passed inspection and faulty devices for defibrillator [22]. This study obtains a perfect accuracy of 100%; however, the study is only limited to a defibrillator.

In addition, Hrvat, et al. [19] used a conformity assessment approach and attained an accuracy of 98.06% to predict a fail and pass response for syringe pumps. Therefore, a new approach is proposed in this research where 44 types of critical medical equipment are utilized using multimodal data to predict three classes. Data analysis saves a cost if the framework is implemented in healthcare services. To the best of our knowledge, this is the first comprehensive model that utilizes 44 types of critical medical equipment with structured and unstructured multimodal data to predict the time to first failure event.

D. STRENGTHEN MAINTENANCE TASKS

By utilizing the LDA techniques, the maintenance task for medical equipment can be strengthened to prevent future failures. The maintenance team shall investigate the components' conditions during routine maintenance. The components' root cause analysis is acquired by plotting the word cloud using the bag of words approach. Every equipment has its unique word cloud for analysis purposes.

The root cause of failures for therapeutic equipment is listed in Table 19. The frequent failure events are listed in the table based on word cloud information depending on the type.

TABLE 20. Root cause of failures for diagnostic and imaging equipment.

No.	Equipment	Failures components
1.	Electrosurgical unit	Diathermy pencil, switch, cable, and paddle faulty
2.	Ultrasound	Probe faulty, printer, cable, power supply, software, image, battery, board, keyboard, and trackball faulty
3.	Apnea monitor	Battery, monitor, keypad and button faulty
4.	Fetal heart detector	Battery, cable, probe, and electronic board faulty
5.	Oxygen monitor	Battery, sensor cell and monitor faulty
6.	Physiologic monitoring	SPO2 probe, sensor, cuff, monitor, cable, battery, hose, wheel, tubing, connector, pump, cable faulty and leaking
7.	Sphygmomanometer	Battery, cuff, connector, hose, leaking, monitor, power supply, cable connector, SPO2 probe, and wheel faulty
8.	Vital sign	Cuff, SPO2 probe and connector, hose, battery, screw, tubing, and adapter faulty
9.	Microscope	Bulb, cable, connection to monitor, and fused faulty
10.	Radiographic, mammographic	System restart, battery, cable connector, module, board, loose connection, and calibration needed
11.	Radiographic, angiographic	Cable, battery, screen, cable connector, software, temperature, and monitor faulty
12.	Computed Tomography	System failure, camera, cable, calibration needed, network, monitor, table faulty, and image artifact
13.	Radiographic, fluoroscopic	Software restart, cable, sensor, table, artifact, wheel, battery, bulb, console keyboard, error, and loose connection
14.	Injector	Heat maintainer, heater, board, calibration, and cable connection
15.	Magnetic Resonance Imaging	System failure, chiller, temperature, network, helium gas refill, connector and cable, monitor, and refrigerator faulty
16.	Radiographic system	Bulb, x-ray tube, screw, cable, calibration, brake, table alignment, and button faulty
17.	Radiographic, dental	Calibration needed, software issues, and loose connection

By knowing these parts, the technical team shall prioritize maintenance works by examining these parts' condition and replacing the components during routine maintenance before failures occur.

Similarly, the root cause of failures for diagnostic and imaging equipment is tabulated in Table 20. Most failure components are related to maintenance wear and tear parts that can be made available beforehand. Thus, such proactive action can reduce waiting time and improve service delivery. This analysis is also beneficial for procurement and parts inventory tracking. The maintenance team can prepare adequate numbers of these components stock at their facility to accommodate the repair work, and parts replacement shall be executed immediately to prevent prolonged downtime.

E. LONG TERM BENEFIT OF THE PROPOSED MODEL

This research introduced the long-term benefit of a comprehensive model to predict the first failure event and enable the industry to prepare resources accordingly. As DL applications have evolved, this study enhances the method with a comparison to ML. The research gap is improved by introducing a failure prediction model comprised of 44 types of critical equipment with a comprehensive cost analysis.

The proposed model is based on thirty years of actual datasets from Government Hospitals in Malaysia. A comprehensive model is more robust than an individual model and will contribute to advancing knowledge in the field. To the best of our knowledge, no analogous study on predicting the performance of medical devices using multimodal data has been conducted since 2010. The conventional way of calculating mean time to failure (MTTF) is executed by dividing the total operating time (hours) by the number of failures [51].

For example, this subsection selects an intensive care ventilator for MTTF calculation. This type is selected due to its criticality, and functionality is essential, especially during outbreaks such as the COVID-19 pandemic. The situation is challenging and worsens when the existing ventilator is malfunctioning.

TABLE 21. MTTF manual calculation for intensive care ventilator.

No.	Device no.	Downtime (hrs)	Failure event (No.)	Operating time based on aged (hrs)	MTTF =Total operating time/ failure event (hrs)
1.	300025128	279.02	11	87,320.98	7,938.27
2.	300025127	155.76	17	87,444.24	5,143.78
3.	300027508	498.88	18	52,061.12	2,892.28
4.	300027510	34.01	11	52,525.99	4,775.09
5.	300027512	372.86	11	52,187.14	4,744.29
6.	300027515	464.23	17	52,095.77	3,064.46
7.	300027514	103.68	15	52,456.32	3,497.09
8.	300027513	256.57	10	52,303.43	5,230.34
9.	300027511	928.45	19	51,631.55	2,717.45
10.	300027509	209.45	11	52,350.55	4,759.14
11.	300025324	32.22	13	87,567.78	6,735.98
12.	300025325	82.38	13	87,517.62	6,732.12
13.	300025326	146.45	12	87,453.55	7,287.80
14.	PRK300/ 000819	2.75	4	61,317.25	15,329.31
15.	PRK358/ 000213	1.25	1	17,518.75	17,518.75

A manual calculation of MTTF is shown in Table 21, where the manual intervention requires two parameters: total operating time and number of failure events. Although a simple analysis could obtain the MTTF result in hours, the calculation did not consider other important elements in every device application.

In addition, MTTF is typically computed assuming constant failure rates and does not consider complex failure modes or degradation patterns. It may not be suitable for systems with non-linear failure behaviors, intermittent failures, or fluctuating operational conditions. AI predictive maintenance can handle more complex scenarios by capturing and analyzing diverse data sources to identify nuanced failure patterns and degradation trends.

Other important parameters in a real industry also influence the time to the failure event. Manual MTTF analysis does not consider other important features contributing to equipment failures, such as age, service support, asset condition, frequency maintenance requirement, maintenance complexity, operations or utilization, maintenance cost, model specification, etc.

MTTF is a reactive maintenance strategy which focuses on failures that have already occurred. It does not provide proactive insights or early warnings about potential malfunctions, which can result in unanticipated delays and expensive repairs. By leveraging machine learning and data analysis, AI predictive maintenance can detect patterns, anomalies, and early failure indications, allowing for timely interventions and preventative measures.

This research offers the application of the stated features in the model development for a comprehensive model framework. The model is able to predict future failure taking into consideration all the necessary parameters. The ML and DL applied in model development can supersede manual calculation and replace it with AI computation.

The LDA is another branch of AI under Natural Language Processing accessible in this study. The technique discovers hidden patterns on the maintenance notes captured during failures, which is impossible using simple analysis. Nevertheless, simple analysis using MTTF may be applicable for small-scale applications. Besides, integrating structured and unstructured data in this work improves the model's accuracy and differs from the current work. The ML model has better accuracy than DL; however, a longer time is required. In addition, a comprehensive cost analysis is proposed to improve the current budget utilization.

Currently, the service contract cost will be incurred as soon as the equipment is purchased, even though it is unnecessary. The study concludes Class 2 in Fig.12 should have a higher maintenance fee rate because it requires more rigorous maintenance than other classes. The service contract fee shall start in the 4th year for all Classes. It is recommended to purchase equipment with three years warranties for Class 2, including routine and correction maintenance; therefore, the service contract cost for the first three years shall be reduced.

In addition, the root cause of components failures for critical equipment is listed based on types to prevent reoccurrence and contribute to better service delivery. Improving the common failures components during routine tasks and maintaining inventory tracking is expected to reduce future failures and shorten the waiting time.

As for Class 3, extended three-year warranties for routine maintenance shall accommodate the actual needs since the equipment is expected to fail after three years. Similarly, service contract costs for the first three years can be reduced. The total approximate cost saving of MYR326,330.88/year is in Fig.13.

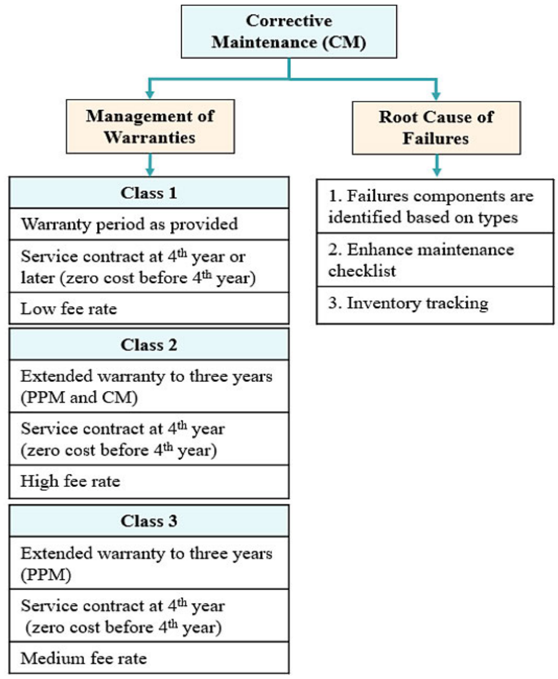

FIGURE 12. Long-term Maintenance Strategies for CM.

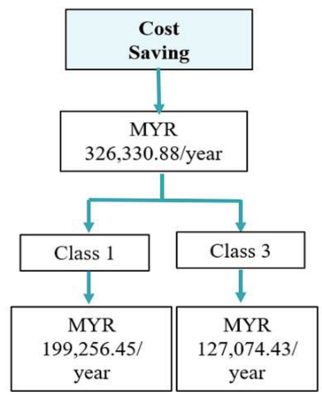

FIGURE 13. Cost Saving Estimation for Classes 1 and 3.

The cost saving for Classes 1 and 3 shall benefit the healthcare system when the model is embarked in the field. The cost saving is obtained by reducing the PPM frequency for Class 1 and Class 3 from bi-annually to annually, as shown in Fig.14. The Class 2 PPM frequency remains and will not involve cost-saving estimation in this paper.

A framework in Fig.15 is proposed for developing a replacement plan policy. In order to ensure that equipment is replaced before the maintenance cost surpasses the purchase cost, a thorough study of the service contract cost and purchase cost is discussed. Based on the likelihood of failure and class severity, Class 2 is prioritized for replacement, followed by Classes 3 and 1.

Three different problems are addressed compared to the present: malfunctioning medical equipment, high maintenance cost and ageing equipment and availability of

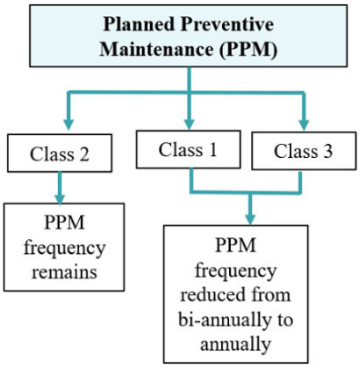

FIGURE 14. Long-term Maintenance Strategies for PPM.

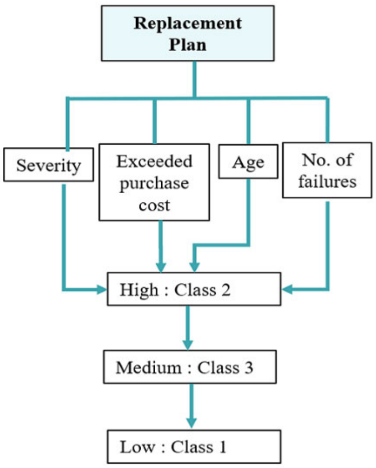

FIGURE 15. Long-term Replacement Plan.

replacement plan policy. Hence, this research contributes to reducing medical equipment failures, decreasing maintenance costs and proposing a replacement plan for hospital management as a long-term strategy.

VI. CONCLUSION

Medical equipment maintenance and replacement are the most expensive aspects of providing healthcare in any country. The proposed predictive model is developed based on multimodal to predict the failure of 8,294 critical medical devices. This paper provides a model to forecast the time until the first failure and classifies Class 1, Class 2, and Class 3. Each paper's section discusses the thorough process, including model optimization and detailed analysis. The final framework is concluded after ML and DL models are compared.

Among all algorithms, Ensemble Classifier performs best after sensitivity analysis with MRMR listed eight significant features, achieving 88.80% accuracy, 94.41% specificity, 88.82% recall, 88.46% precision, and 88.84% F1 Score. Instead of using simply structured data, LDA capacity to discover failure hidden patterns increases the model's potential and dependability. The analysis is limited by data extracted directly from ASIS and hospital user failure complaints. There is a possibility that the maintenance system does not record verbal complaints or events that have been concealed.

The upcoming work will investigate and estimate the timing of the subsequent occurrence. The prediction must include the time between the first and second failure and the time-lapse. The ensuing failures study may suggest equipment life cycle analysis and remaining useful life for medical equipment.

ACKNOWLEDGMENT

The authors would like to express their highest appreciation to the Director General of Health Malaysia for the medical device dataset of healthcare facilities in Peninsular Malaysia.

REFERENCES

[1] M. Masmoudi, Z. B. Houria, A. Al Hanbali, and F. Masmoudi, "Decision support procedure for medical equipment maintenance management," *J. Clin. Eng.*, vol. 41, no. 1, pp. 19–29, Jan. 2016, doi: 10.1097/JCE.0000000000000135.

[2] *Medical Device Act*, Department of Standards Malaysia, Cyberjaya, Malaysia, 2012.

[3] J. K. Kutor, P. Agede, and R. H. Ali, "Maintenance practice, causes of failure and risk assessment of diagnostic medical equipment," *J. Biomed. Eng. Med. Devices*, vol. 2, no. 1, p. 123, 2017, doi: 10.4172/2475-7586.1000123.

[4] H. Mahfoud, E. B. Abdellah, and A. El Biyaali, "Dependability-based maintenance optimization in healthcare domain," *J. Quality Maintenance Eng.*, vol. 24, no. 2, pp. 200–223, May 2018, doi: 10.1108/JQME-07-2016-0029.

[5] B. S. Dhillon, "Medical equipment reliability: A review, analysis methods and improvement strategies," *Int. J. Rel., Quality Saf. Eng.*, vol. 18, no. 4, pp. 391–403, Aug. 2011.

[6] J. Palmer, "Critical incidents: The respiratory system," *Anaesthesia Intensive Care Med.*, vol. 11, no. 10, pp. 403–406, Oct. 2010.

[7] C. Verbano and F. Turra, "A human factors and reliability approach to clinical risk management: Evidence from Italian cases," *Saf. Sci.*, vol. 48, no. 5, pp. 625–639, Jun. 2010, doi: 10.1016/j.ssci.2010.01.014.

[8] R. Thapa, A. Yih, A. Chauhan, S. Poudel, S. Singh, S. Shrestha, S. Tamang, R. Shrestha, and R. Rajbhandari, "Effect of deploying biomedical equipment technician on the functionality of medical equipment in the government hospitals of rural Nepal," *Human Resour. Health*, vol. 20, no. 1, p. 21, Dec. 2022, doi: 10.1186/s12960-022-00719-y.

[9] Grand View Research, "Medical equipment maintenance market size, share & trends analysis report by equipment (imaging equipment, surgical instruments), by service (corrective maintenance, preventive maintenance), and segment forecasts, 2021–2027," Grand View Res., San Francisco, CA, USA, Tech. Rep. GVR-4-68039-308-0, Jul. 2022.

[10] Markets and Markets. *Medical Equipment Maintenance Market by Product (Imaging (MRI, CT, PET-CT, Ultrasound, X-Ray), Endoscopes, Lasers, Ventilators, Dialysis, Monitors), Provider (OEM, ISO, In-House), Service (Preventive, Operational), User (Hospital, ASCs)—Forecast to 2026*. Accessed: Dec. 3, 2021. [Online]. Available: https://www.marketsandmarkets.com

[11] W. Suraya and W. M. Radzi. (2021). *Auditor's General Report*. [Online]. Available: https://www.audit.gov.my

[12] B. Raju, F. Jumah, O. Ashraf, V. Narayan, and G. Gupta, "Big data, machine learning, and artificial intelligence: A field guide for neurosurgeons," *J. Neurosurgery*, vol. 135, no. 2, pp. 373–383, Aug. 2021, doi: 10.3171/2020.5.JNS201288.

[13] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2 Vec," *Inf. Sci.*, vol. 477, pp. 15–29, Mar. 2019, doi: 10.1016/j.ins.2018.10.006.

[14] C.-N. Wang, F.-C. Yang, N. T. M. Vo, and V. T. T. Nguyen, "Wireless communications for data security: Efficiency assessment of cybersecurity industry—A promising application for UAVs," *Drones*, vol. 6, no. 11, p. 363, Nov. 2022, doi: 10.3390/drones6110363.

[15] D.-E.-M. Nisar, R. Amin, N.-U.-H. Shah, M. A. A. Ghamdi, S. H. Almotiri, and M. Alruily, "Healthcare techniques through deep learning: Issues, challenges and opportunities," *IEEE Access*, vol. 9, pp. 98523–98541, 2021, doi: 10.1109/ACCESS.2021.3095312.

[16] A. Jamshidi, S. A. Rahimi, D. Ait-kadi, and A. Ruiz, "A comprehensive fuzzy risk-based maintenance framework for prioritization of medical devices," *Appl. Soft Comput.*, vol. 32, pp. 322–334, Jul. 2015, doi: 10.1016/j.asoc.2015.03.054.

[17] M. Arathy and K. Balasubramanian, "Fuzzy logic based failure modes and effects analysis on medical ventilators," in *Proc. 5th Int. Conf. Commun. Electron. Syst. (ICCES)*, Jun. 2020, pp. 962–966.

[18] A. O. Hutagalung and S. Hasibuan, "Determining the priority of medical equipment maintenance with analytical hierarchy process," *Int. J. Online Biomed. Eng.*, vol. 15, no. 10, pp. 107–120, 2019, doi: 10.3991/ijoe.v15i10.10920.

[19] F. Hrvat, L. Spahic, L. G. Pokvic, and A. Badnjevic, "Artificial neural networks for prediction of medical device performance based on conformity assessment data: Infusion and perfusor pumps case study," in *Proc. 9th Medit. Conf. Embedded Comput. (MECO)*, Jun. 2020, pp. 1–4.

[20] L. Spahić, "Machine learning techniques for performance prediction of medical devices: Infant incubators," in *Proc. CMBEBIH*, A. Badnjevic, R. Škrbić, and L. G. Pokvić, Eds. Cham, Switzerland: Springer, 2020, pp. 483–490.

[21] Ž. Kovačević, L. G. Pokvic, L. Spahić, and A. Badnjević, "Prediction of medical device performance using machine learning techniques: Infant incubator case study," *Health Technol.*, vol. 10, no. 1, pp. 151–155, Jan. 2020, doi: 10.1007/s12553-019-00386-5.

[22] A. Badnjević, L. G. Pokvić, M. Hasičić, L. Bandić, Z. Mašetić, Ž. Kovačević, J. Kevrić, and L. Pecchia, "Evidence-based clinical engineering: Machine learning algorithms for prediction of defibrillator performance," *Biomed. Signal Process. Control*, vol. 54, Sep. 2019, Art. no. 101629, doi: 10.1016/j.bspc.2019.101629.

[23] R. Bahreini, L. Doshmangir, and A. Imani, "Influential factors on medical equipment maintenance management," *J. Quality Maintenance Eng.*, vol. 25, no. 1, pp. 128–143, Mar. 2019, doi: 10.1108/JQME-11-2017-0082.

[24] A. H. Zamzam, A. K. I. Al-Ani, A. K. A. Wahab, K. W. Lai, S. C. Satapathy, A. Khalil, M. M. Azizan, and K. Hasikin, "Prioritisation assessment and robust predictive system for medical equipment: A comprehensive strategic maintenance management," *Frontiers Public Health*, vol. 9, p. 1805, Nov. 2021, doi: 10.3389/fpubh.2021.782203.

[25] H.-Y. Liao, W. Cade, and S. Behdad, "Forecasting repair and maintenance services of medical devices using support vector machine," *J. Manuf. Sci. Eng.*, vol. 144, no. 3, Mar. 2022, Art. no. 031005, doi: 10.1115/1.4051886.

[26] M. Xia, T. Li, T. Shu, J. Wan, C. W. de Silva, and Z. Wang, "A two-stage approach for the remaining useful life prediction of bearings using deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 6, pp. 3703–3711, Jun. 2019, doi: 10.1109/TII.2018.2868687.

[27] K. T. P. Nguyen and K. Medjaher, "A new dynamic predictive maintenance framework using deep learning for failure prognostics," *Rel. Eng. Syst. Saf.*, vol. 188, pp. 251–262, Aug. 2019, doi: 10.1016/j.ress.2019.03.018.

[28] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, Jan. 2018, doi: 10.1016/j.neucom.2017.07.032.

[29] B. Chen, Y. Liu, C. Zhang, and Z. Wang, "Time series data for equipment reliability analysis with deep learning," *IEEE Access*, vol. 8, pp. 105484–105493, 2020, doi: 10.1109/ACCESS.2020.3000006.

[30] *Topographic-Map.Com*. Accessed: Jun. 25, 2022. [Online]. Available: https://en-us.topographic-map.com/

[31] J. Wang, X. Sun, M. F. H. She, A. Kouzani, and S. Nahavandi, "Unsupervised mining of long time series based on latent topic model," *Neurocomputing*, vol. 103, pp. 93–103, Mar. 2013, doi: 10.1016/j.neucom.2012.09.008.

[32] C. Karmakar and M. Datcu, "A framework for interactive visual interpretation of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3161959.

[33] I. Tabian, H. Fu, and Z. S. Khodaei, "A convolutional neural network for impact detection and characterization of complex composite structures," *Sensors*, vol. 19, no. 22, p. 4933, Nov. 2019, doi: 10.3390/s19224933.

[34] S. Sultana, S. S. Hussain, M. Hashmani, J. Ahmad, and M. Zubair, "A deep learning hybrid ensemble fusion for chest radiograph classification," *Neural Netw. World*, vol. 31, no. 3, pp. 191–209, 2021, doi: 10.14311/NNW.2021.31.010.

[35] T. Kareen, "Deep learning on electronic medical records for hospital readmission risk prediction," Doctor Philosophy, Fac. Eng., Univ. Malaya, Lumpur, Malaysia, 2020.

[36] J. L. Garcia-Balboa, M. V. Alba-Fernandez, F. J. Ariza-López, and J. Rodriguez-Avi, "Homogeneity test for confusion matrices: A method and an example," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 1203–1205, doi: 10.1109/IGARSS.2018.8517924.

[37] O. Pastor-Bárcenas, E. Soria-Olivas, J. D. Martín-Guerrero, G. Camps-Valls, J. L. Carrasco-Rodríguez, and S. D. Valle-Tascón, "Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling," *Ecological Model.*, vol. 182, no. 2, pp. 149–158, Mar. 2005, doi: 10.1016/j.ecolmodel.2004.07.015.

[38] N. M. Gazzaz, M. K. Yusoff, A. Z. Aris, H. Juahir, and M. F. Ramli, "Artificial neural network modeling of the water quality index for Kinta river (Malaysia) using water quality variables as predictors," *Mar. Pollut. Bull.*, vol. 64, no. 11, pp. 2409–2420, Nov. 2012, doi: 10.1016/j.marpolbul.2012.08.005.

[39] S. Chen, Y. Ren, D. Friedrich, Z. Yu, and J. Yu, "Sensitivity analysis to reduce duplicated features in ANN training for district heat demand prediction," *Energy AI*, vol. 2, Nov. 2020, Art. no. 100028, doi: 10.1016/j.egyai.2020.100028.

[40] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proc. Comput. Syst. Bioinformatics IEEE Bioinf. Conf.*, Aug. 2003, pp. 523–528, doi: 10.1109/CSB.2003.1227396.

[41] A. Anuragi, D. Singh Sisodia, and R. B. Pachori, "Epileptic-seizure classification using phase-space representation of FBSE-EWT based EEG sub-band signals and ensemble learners," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103138, doi: 10.1016/j.bspc.2021.103138.

[42] S. Al-Juboori, "Medical devices service life cycle cost management in Al Karak hospital as a case study," *J. Accounting Marketing*, vol. 4, no. 2, p. 134, 2015, doi: 10.4172/2168-9601.1000134.

[43] G. Bektemur, N. Muzoğlu, M. A. Arıcı, and M. K. Karaaslan, "Cost analysis of medical device spare parts," *Pakistan J. Med. Sci.*, vol. 34, no. 2, pp. 472–477, Apr. 2018, doi: 10.12669/pjms.342.14245.

[44] M. Hillebrecht, C. Schmidt, B. P. Saptoka, J. Riha, M. Nachtnebel, and T. Bärnighausen, "Maintenance versus replacement of medical equipment: A cost-minimization analysis among district hospitals in Nepal," *BMC Health Services Res.*, vol. 22, no. 1, p. 1023, Aug. 2022, doi: 10.1186/s12913-022-08392-6.

[45] D. S. Kumar, *Medical Equipment Maintenance Policy*. New Delhi, India: Employees' State Insurance Corporation, 2012.

[46] T. W. Gamessa, S. T. Abebe, L. D. Abate, M. K. Abo, A. A. Mekonnen, Z. K. Tadesse, A. F. Woyesa, R. B. Obse, M. A. Ibrahim, and G. Simegn, "Planning and budgeting of medical devices among Ethiopian public hospitals," *ClinicoEconomics Outcomes Res.*, vol. Volume 14, pp. 405–413, May 2022, doi: 10.2147/ceor.S363376.

[47] A. Jamshidi, S. A. Rahimi, D. Ait-Kadi, and A. Ruiz, "Medical devices inspection and maintenance; a literature review," in *Proc. IIE Annu. Conf.*, 2014, pp. 3895–3904.

[48] B. Wang, T. Rui, and S. Balar, "An estimate of patient incidents caused by medical equipment maintenance omissions," *Biomed. Instrum. Technol.*, vol. 47, no. 1, pp. 84–91, Jan. 2013, doi: 10.2345/0899-8205-47.1.84.

[49] *Medical Equipment Maintenance Programme Overview*, World Health Org., Geneva, Switzerland, 2011.

[50] C. Stenström, P. Norrbin, A. Parida, and U. Kumar, "Preventive and corrective maintenance–cost comparison and cost–benefit analysis," *Struct. Infrastructure Eng.*, vol. 12, no. 5, pp. 603–617, May 2016, doi: 10.1080/15732479.2015.1032983.

[51] P. Kosky, R. Balmer, W. Keat, and G. Wise, "Chapter 11—Industrial engineering," in *Exploring Engineering*, 5th ed. New York, NY, USA: Academic, 2021, pp. 229–257.

NOORUL HUSNA ABD RAHMAN received the B.Eng. degree in electrical electronic engineering from Universiti Tenaga Nasional (UNITEN), in 2013. She has been a Biomedical Engineer with the Engineering Services Division, Ministry of Health, Malaysia, since 2014. Her scope of work is on the privatization of hospital support services, operation, and maintenance especially in biomedical engineering. She provides technical advice on medical device technical specifications and system installation for healthcare facilities under the Ministry of Health. On professional qualification, she is a Professional Engineer (P.Eng.) in electrical engineering discipline registered with the Board of Engineers Malaysia (BEM), a Corporate Member of IEM, an ASEAN Certified Energy Manager (AEMAS), a Certified Qualified Professional (MyCREST Green Certification), and a Certified Radiation Protection Officer registered with the Atomic Energy Licensing Board, Malaysia.

KHAIRUNNISA HASIKIN received the B.Eng. degree in electrical engineering and the M.Sc. Eng. degree in biomedical engineering from the University of Malaya, Malaysia, and the Ph.D. degree from Universiti Sains Malaysia. She is currently a Senior Lecturer with the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya. On professional qualification, she is a Professional Engineer (P.Eng.) in electrical engineering discipline registered with the Board of Engineers Malaysia (BEM). She is actively engaging herself in research. Her main research interests include artificial intelligence (AI), machine learning, and image processing in various fields with a specific interest in medical/healthcare applications. She is also channeling her expertise in AI for sustainability management projects specifically in discovering novel pollutant markers for intelligent environmental decision systems.

NASRUL ANUAR ABD RAZAK received the B.Eng. degree in mechatronics from International Islamic University Malaysia, Selangor, Malaysia, in 2009, and the M.Eng. and Ph.D. degrees from the Department of Biomedical Engineering, University of Malaya, Kuala Lumpur, Malaysia, in 2011 and 2014, respectively. Since 2015, he has been a Senior Lecturer with the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya. His research interests include healthcare technology management, assistive device management, biomechanical engineering (manufacturing in prosthetics and bio sculptor CAD/CAM systems), prosthetics and orthotics engineering, and rehabilitation engineering (upper limb prosthetic, assistive devices and technology in prosthetic and orthotics, and biomechatronic in prosthetics and orthotics).

AYMAN KHALLEL AL-ANI received the Ph.D. degree in advance computer network from Universiti Sains Malaysia (USM). He is currently a Senior Lecturer with the Faculty of Computing and Informatics, Universiti Malaysia Sabah (UMS). His current research interests include malware detection, web security, intrusion detection systems (IDS), intrusion prevention systems (IPS), network monitoring, the Internet of Things (IoT), IPv6 security, artificial intelligence, machine learning, data mining, and optimization algorithms.

D. JERLINE SHEEBHA ANNI received the B.Tech. degree in information technology and the Master of Engineering degree in computer science and engineering from Anna University, Chennai, India, and the Ph.D. degree from VIT University, India. She is currently a Principal and a Professor with the Department of Computer Science and Engineering, KMCT College of Engineering for Women, Kerala, India. Her research interests include wireless sensor networks, image processing, sound classification, and medical image processing.

PRABU MOHANDAS received the B.Tech. degree in information technology and the Master of Engineering degree in computer science and engineering from Anna University, Chennai, India, and the Ph.D. degree from Anna University. He is currently an Assistant Professor with the Department of Computer Science and Engineering, National Institute of Technology Calicut, Kerala, India. His research interests include wireless sensor networks, machine learning, elephant movement pattern prediction, and (mammograms) medical image processing.