Res Intell Manuf Assem, 2025, 4(2): 219-238
DOI: 10.25082/RIMA.2025.02.001

RESEARCH ARTICLE

Machine Learning Approaches to Predicting Pacemaker Battery Life

Samikshya Neupane[1] **Tarun Goswami**[1*]

[1] Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, Dayton, OH 45435, USA

Correspondence to: Tarun Goswami, Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, Dayton, OH 45435, USA; Email: tarun.goswami@wright.edu

Received: April 14, 2025;
Accepted: June 20, 2025;
Published: June 25, 2025.

Citation: Neupane S, Goswami T. Machine Learning Approaches to Predicting Pacemaker Battery Life. *Res Intell Manuf Assem*, 2025, **4**(2): 219-238.
https://doi.org/10.25082/RIMA.2025.02.001

Abstract: Accurate prediction of pacemaker battery life is critical to timely generator replacement and patient safety. We evaluated three regression approaches: multilayer perceptron Neural Networks (NN), Random Forests (RF), and Linear Regression (LR), using 42 real-world interrogation reports spanning single, dual, and triple-chamber Medtronic devices. Key electrical parameters (battery voltage/current, lead impedance, capture thresholds, pacing percentages, *etc.*) were modelled. Performance was quantified with mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R^2). NNs achieved the highest accuracy ($R^2 \approx 1.0$; MAE < 0.1 months), RF provided robust results ($R^2 \approx 0.85$), whereas LR exhibited limited predictive fidelity ($R^2 \leq 0.41$). "Monte-Carlo simulations (n = 1000)" and 95% prediction intervals characterized predictive uncertainty; residual and Q-Q analyses verified statistical assumptions. Our findings indicate that a data-driven NN framework can reliably forecast remaining battery longevity, enabling proactive replacement scheduling and reducing unexpected generator depletion. The methodology is compatible with different manufacturers and suitable to integration within remote device follow-up systems to enhance longitudinal cardiac care.

Keywords: battery life prediction, machine learning models, neural networks, random forests regression, linear regression, Monte Carlo simulations

1 Introduction

Arrhythmia or irregular heartbeats are major contribution to cardiovascular diseases, a group which continues to feature prominently both when it comes to morbidity and mortality globally [1]. Pacemakers are extraordinarily successful tool for handling such irregularities as they effectively restore and preserve normal heart rhythms [2]. For millions of patients worldwide, these small, implanted devices greatly enhance their quality of life by delivering electrical impulses that control the heart's rhythm [3]. Since their inception in the 1950s, pacemakers have evolved significantly, enabling adaptive pacing, continuous monitoring, and personalized therapies [4,5].

The modalities of modern pacemakers are categorized to almost types based on the chambers they pace and the way they work. There are single-chamber pacemakers (stimulates atrium or ventricle) and dual-chamber pacemakers (stimulates both chambers in a coordinated manner. Newer models, like biventricular pacemakers, can be used in heart failure patients that will pace into both ventricles and improve efficiency of the heart. These advancements have made pacemakers an indispensable device in the treatment and management of several cardiovascular conditions, ranging from bradycardia to complex arrhythmias [6].

1.1 Single-chamber Pacemaker

The single-chamber pacemaker is a type of pacemaker which has a single chamber, monitors the heart's rhythm and sends electrical impulses when it detects an abnormal or slow heartbeat. Single-chamber pacemakers are effective at pacing the heart at the proper rate, but they can't manage more sophisticated arrhythmias that impact several heart chambers. For bradycardia, where the heart beats too slowly, patients are often recommended this device and for the patients requiring pacing in only one chamber of the heart usually the right ventricle [7]. (Figure 1)

1.2 Dual-Chamber Pacemaker

Both right atrium and right ventricle are stimulated in a dual-chamber pacemaker which supports more sophisticated pacing. This pacemaker is most suited for patients with atrioven-

Figure 1 Diagram of a single-chamber pacemaker [8]

tricular (AV) block or other pathologies that occur downstream of communication between atria and ventricles. By pacing both chambers, the dual-chamber pacemaker promotes correct timing contrast between atrial and ventricular contractions, improving overall efficiency of the heart and preventing pathology such as atrial fibrillation or congestive heart failure. Compared to single-chamber devices, this kind of pacemaker provides better synchronization between the heart's chambers, leading to more natural cardiac rhythms and better clinical outcomes [9]. (Figure 2)

Figure 2 Diagram of a dual-chamber pacemaker [10]

1.3 Triple-Chamber Pacemaker

Also known as a cardiac resynchronization treatment (CRT) device, a triple-chamber pacemaker is the most sophisticated of all pacemaker types. It is used to help patients who have very poor heart function or other conditions that result in poor blood circulation into the heart. This was very different from the traditional pacemaker with two different chambers. Resynchronization of the electrical activity of the heart improves the pump function of the heart which in turn improves the heart failure symptoms, and improves the quality of life of the patient. CRT plays an important role in left ventricle dysfunction when other modalities fail [11]. (Figure 3)

Figure 3 Diagram of a triple-chamber pacemaker [12]

Lead impedance, pacing thresholds, device programming, and the particular cardiac condition being treated are some of the variables that affect these pacemaker's longevity and performance [13]. The requirement for precise battery life and device performance monitoring increases as pacemakers get more complicated, moving from single-chamber to dual-chamber and triple-chamber devices. Optimizing patient care and device maintenance requires an understanding of how these various pacemaker types work as well as the variables that affect how long their batteries operate.

1.4 Battery technology in Medtronic Pacemakers

Pacemaker battery longevity is influenced by a combination of device type, pacing behavior, and individual patient factors. In general, single-chamber pacemakers tend to have longer battery life due to lower energy demands, while dual- and triple-chamber devices consume more energy as they are required to pace multiple cardiac chambers and maintain synchronized contractions. The majority of modern pacemakers use lithium-based batteries, with lithium-iodine (Li-I_2) cells being a common choice across manufacturers due to their high energy density, long-term stability, and biocompatibility. These batteries form a solid electrolyte, minimizing leakage risks and making them ideal for long-term implantation. Battery life can range from 10–15 years in single-chamber devices, 8–12 years in dual-chamber, and 5–10 years in triple-chamber pacemakers, although actual lifespan varies based on lead impedance, capture thresholds, pacing percentage, and patient activity levels. Manufacturers often implement energy management systems to optimize power usage and extend device longevity. For example, Medtronic, a leading pacemaker manufacturer, utilizes lithium-iodine batteries and incorporates advanced energy-saving algorithms in its devices. While Medtronic devices were used in this study, the modeling approach and findings are designed to be applicable across all pacemaker platforms that use similar battery and pacing technologies [14–18].

1.5 Importance of Battery Life Prediction

Ensuring pacemaker functionality and longevity depends on effective battery management. That means we need to watch battery usage carefully. Over a long time, each pacemaker battery gradually loses power, we change them on time to avoid bad functioning of its device. Data about how much power the battery has left is very essential for both patient care and planning replacement time as early failure could cause big trouble [19].

This research is focused on using machine learning models for pacemaker, especially for forecasting battery life and device performance. These algorithms can forecast the battery's lifespan and foresee possible failures by examining data from pacemaker interrogation reports, giving clinicians important information. These prediction models find trends in pacemaker behavior, including pacing thresholds, lead impedance, and battery condition, by using data mining techniques, neural networks, and other machine learning algorithms. The application of these models in clinical settings has enormous potential to increase pacemaker monitoring accuracy and improve patient outcomes in general [20].

The varieties of pacemakers, their battery life issues, and the use of machine learning models to enhance pacemaker management are the main topics of this research for the development of cardiovascular devices. The study lays the groundwork for comprehending the intricacies of pacemaker function, the function of predictive models, and the necessity of ongoing pacemaker technology improvement by giving a broad review of these subjects.

1.6 Data Mining

Data mining, a potent technology that combines methods from statistics, machine learning, and artificial intelligence, is used to identify important patterns in large and complex data sets, especially in sectors like healthcare, finance, and biomedical engineering [21]. Biomedical engineering greatly depends on the analysis of complex medical data, such as diagnostic imaging, medical records, and device reports, to improve patient care and device management [22]. Predicting pacemaker battery life is one significant use, which is essential for guaranteeing patient safety and device dependability [23]. Battery voltage, lead impedance, and pacing thresholds are among the data included in pacemaker interrogation reports yet, conventional techniques frequently fail to recognize their intricate relationships [24]. Neural networks, random forests, and linear regression are examples of machine learning models that reveal hidden patterns [18, 21]. Building such predictive models is the main goal of this research in order to assist medical professionals in improving patient outcomes, cutting down on pointless treatments, and making better decisions.

1.7 Machine Learning Models for Pacemaker Battery Life Prediction

Machine learning approaches like neural networks, random forests, and linear regression provide strong instruments for examining intricate datasets and identifying patterns that conventional approaches can miss in the context of pacemaker battery life prediction. These types are all well-suited to handle the complex nature of pacemaker battery performance since they each offer unique benefits. This work aims to improve the accuracy of battery life estimates and dependability by applying these models to pacemaker interrogation reports, giving doctors a more reliable and data-driven approach to device care.

1.7.1 Neural Networks (NN)

A class of machine learning models called neural networks draws inspiration from the composition and operations of the human brain. These models are perfect for predicting pacemaker battery life because of their exceptional ability to handle big, complex datasets and non-linear connections. Interconnected nodes, or neurons, arranged in input, hidden, and output layers make up a neural network. Neural networks are capable of capturing intricate relationships between many parameters, including lead impedance, pacing settings, and device usage, in the context of pacemaker data. They are especially useful for producing individualized predictions and raising the precision of battery life estimations because of their capacity to adjust to hidden patterns and simulate complex relationships between factors [25].

1.7.2 Random Forest (RF)

An ensemble learning method called Random Forest builds several decision trees by choosing subsets of the data and features at random. Random forests can generate predictions based on the majority vote of all the trees in the ensemble because of this diversity. The ability of random forests to withstand overfitting is one of their main advantages, which makes them ideal for managing the diverse and noisy nature of pacemaker data. The most important elements influencing battery lifespan, such as lead impedance and pacing demand, can be found using random forests in the context of battery life prediction. Additionally, random forests offer a feature relevance metric that can help physicians prioritize the most important aspects of pacemaker device management [26].

1.7.3 Linear Regression (LR)

A popular statistical model that makes the assumption that there is a linear relationship between the input features and the target variable is called linear regression. When it comes to pacemaker battery life prediction, linear regression provides an easy-to-understand method for figuring out how specific elements, such battery voltage or pacing threshold, affect battery depletion. Linear regression offers a helpful starting point for locating linear dependencies in the data, but being less adaptable than more sophisticated models like neural networks. It is a useful tool for preliminary study and for comprehending the overall impact of certain parameters on pacemaker battery performance because of its simplicity and transparency [27].

This research seeks to create a thorough understanding of pacemaker battery life prediction by utilizing the strengths of these three different models: random forest's resilience in managing a variety of data, neural network's ability to adapt to complex, non-linear relationships, and linear regression's ease of use and interpretability. By combining these models, it is possible to analyze pacemaker interrogation data in a comprehensive way, increase the precision of battery life estimates, and ultimately improve patient care and device management.

1.8 Monte Carlo Simulation in Predictive Modeling

Monte Carlo simulation has become a potent method for evaluating prediction uncertainty and variability in order to further improve the dependability of predictive models. A computing method called Monte Carlo simulation models the probability distribution of outcomes in systems with inherent uncertainty by repeatedly sampling at random. By adding small changes to input data, like pacing thresholds, lead impedance, and battery voltage, Monte Carlo simulation enables researchers to model a variety of potential situations in the context of pacemaker battery life prediction [28]. Together with confidence intervals that measure the degree of uncertainty in the forecasts, this method offers a probabilistic estimate of battery life.

Because real-world data frequently contains noise and variability, Monte Carlo simulation is very useful in healthcare contexts. Clinicians can make better decisions regarding device maintenance and replacement by using Monte Carlo methods to establish the range of probable battery life forecasts by simulating thousands of different events [29]. For instance, a tight Monte Carlo simulation confidence interval denotes high confidence in the anticipated battery life, whereas a broader range denotes increased uncertainty and calls for closer device monitoring [30].

The robustness of machine learning methods, such as neural networks, random forests, and linear regression, in forecasting pacemaker battery life is assessed in this study using Monte Carlo simulation. Through the use of Monte Carlo techniques, this study seeks to improve the accuracy of forecasts for clinical decision-making and offer a more thorough understanding of the variables affecting battery longevity.

1.9 Analytical Tools for Model Evaluation

This study used a number of analytical approaches to evaluate the effectiveness and dependability of predictive models. In order to verify that the prediction errors adhere to a normal distribution, a crucial premise in regression modeling: normal probability plots, or Q-Q plots, were employed to assess the residual's normality. Patterns in prediction mistakes were found by analyzing residual plots, which may indicate biases or systematic departures in the models. Confidence intervals, which provide a range within which the actual battery life is anticipated to fall with a given likelihood (*e.g.*, 95%), were also computed to quantify the uncertainty in forecasts. These resources were utilized in conjunction with Monte Carlo simulations to evaluate the model's resilience and offer a thorough comprehension of their capacity for prediction [31].

In addition to these tools, Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) are important metrics used to assess how well the machine learning models perform.

1.9.1 Mean Absolute Error (MAE)

The average absolute difference between the expected and actual values is measured by MAE. Since it shows the average difference between predictions and actual values, it offers a clear explanation of the model's error size. Lower MAE values indicate better model performance, as they reflect smaller prediction errors [32, 33].

1.9.2 Mean Squared Error (MSE)

The average of the squared discrepancies between expected and actual data is determined by MSE. MSE is sensitive to outliers since it penalizes greater errors more severely by squaring the errors. Because it represents reduced total prediction errors, a lower MSE denotes a more accurate model [34].

1.9.3 R-squared (R^2)

The coefficient of determination, which we often just refer to as R Squared, measures how much variation in dependent variable can be explained by the independent variables. Its value ranges from zero to one; zero means we can predict nothing, and one means the model precisely explains all variation in data. A better match between the model and the data is shown by higher R^2 values [34].

These indicators are crucial for evaluating how well prediction models, like the ones employed in this work to estimate pacemaker battery life perform. We may assess the models' accuracy, precision, and dependability and make sure they are appropriate for clinical decision-making by examining MAE, MSE, and R^2.

1.10 Research objective

The goal of this study is to develop and evaluate predictive models for estimating the remaining battery life of pacemakers using real-world interrogation data. This work focuses on applying machine learning algorithms including Neural Networks, Random Forests, and Linear Regression to forecast battery life based on interrogation data from a range of pacemaker types (single-, dual-, and triple-chamber). Rather than emphasizing a specific manufacturer, this study explores a generalizable framework applicable across devices, reflecting the shared challenges in battery life prediction across the industry. The sub-goals of the research include:

(1) Identifying key parameters from interrogation reports that influence battery depletion, such as pacing thresholds, lead impedance, and voltage trends.

(2) Training and validating multiple machine learning models on these parameters to assess prediction accuracy using performance metrics such as MAE, MSE, and R^2.

(3) Quantifying model uncertainty through Monte Carlo simulations and confidence interval analysis, ensuring robustness under real-world variability.

(4) Evaluating clinical usability of the models by analyzing prediction distributions and determining whether predictions can support timely and cost-effective device replacement.

2 Materials and Methods

This section outlines the methodology employed to predict pacemaker battery life using advanced data mining techniques and Python. The study involves analyzing data from interrogated pacemakers of three types: single-chamber, dual-chamber, and triple-chamber pacemakers. Machine learning models were trained to predict the remaining battery life based on extracted features. Data preprocessing, feature selection, model training, evaluation, and validation processes are described in detail.

2.1 Materials

The primary material used for this study includes pacemaker interrogation reports obtained from our laboratory. The pacemakers used in the study were obtained from the anatomical gift program, meaning posthumously, therefore, no ethical approval was waived. All pacemakers were in vivo for a period of time, and not tested in any tank.

The data set contains interrogation reports from different types of Medtronic pacemakers (models listed in Appendix A), and we categorize these into three groups: single-chamber, dual-chamber and triple-chamber. We have a total of 42 reports, with 14 of them stemming from single-chamber pacemakers, 21 from dual-chamber and 7 from triple-chamber pacemakers. The remaining battery life was extracted directly from the device interrogation reports. These values represent internal device estimates, not independently verified ground truth or stress-test data. As such, the model predicts the remaining battery life on the basis of trained data sets (80%) and compares to the manufacturer-calculated values for the tested values (20%). The key components of the materials used in this study include:

(1) Pacemaker Data: Interrogation reports from Medtronic pacemakers, containing variables such as battery voltage, battery impedance, lead impedance, battery current, pacing percentage, capture threshold, and pulse width.

(2) Software Tools: Python for data preprocessing, model development, statistical analysis, additional data handling and machine learning model implementation.

(3) Computational Resources: A high-performance workstation with at least 16GB RAM executing Monte Carlo simulations and training machine learning models.

2.2 Data Collection and Preprocessing

2.2.1 Dataset

The dataset comprises interrogation reports of Medtronic pacemaker. The data were categorized into three groups based on the pacemaker type:

(1) Single-chamber pacemakers: 14 interrogation reports;

(2) Dual-chamber pacemakers: 21 interrogation reports;

(3) Triple-chamber pacemakers: 7 interrogation reports.

Each category has distinct features relevant to its chamber configuration. While we acknowledge that the triple-chamber pacemaker group contains only 7 samples, we elected to keep it separate from the dual-chamber group due to its distinct functional profile and unique feature set (*e.g.*, additional lead impedance values, resynchronization pacing parameters). Triple-chamber devices have substantially higher energy consumption patterns and distinct programming logic compared to dual-chamber pacemakers. Integrating the two would risk blending functionally distinct behaviors, compromising the interpretability and clinical utility of the models.

2.2.2 Data Preprocessing and Handling Missing Values

Before applying machine learning models, the data underwent preprocessing steps to ensure its quality and integrity:

(1) Data Preprocessing: Identifying and handling missing values using imputation techniques (mean/mode replacement and predictive imputation).

(2) Data Normalization: Standardizing numerical features using min-max scaling for better model convergence.

(3) Outlier Detection: Identifying and removing extreme values.

(4) Feature Encoding: Converting categorical variables (if any) into numerical form.

2.2.3 Feature Engineering

Relevant features were selected based on their impact on pacemaker battery life prediction. Each pacemaker type had different feature sets due to variations in lead configurations and sensor readings. For instance, in the case of triple-chamber pacemakers, the selected features included:

(1) Lead impedance values;

(2) Capture thresholds;

(3) Programmed amplitudes and pulse widths;

(4) Measured P/R waves;

(5) Programmed sensitivity values;

(7) Lower and upper sensor rate;

(8) Pacing percentages (sensed and paced);

(9) Patient activity and time in AT/AF per day.

Features for dual-chamber pacemakers:

(1) Lead impedance;

(2) Battery Voltage;
(3) Capture Threshold;
(4) Programmed Amplitude;
(5) Pulse Width;
(6) Measured P/R wave;
(7) Programmed sensitivity;
(8) Lower sensing Rate;
(9) Pacing Percentage (Sensed and Paced);
(10) Patient activity hr/day;
(11) Time in AT/AF hr/day.

Features for single-chamber pacemakers:
(1) Battery Current;
(2) Lead Current;
(3) Battery impedance;
(4) Battery Voltage;
(5) Capture Threshold;
(6) Programmed Amplitude;
(7) Pulse Width;
(8) Pacing Percentage (Sensed and Paced).

2.2.4 Data Scaling

To ensure uniformity in the feature space, standardization was applied using "StandardScaler", transforming the data into a standardized format.

2.3 Machine learning Models

Three regression models were employed to predict pacemaker battery life:
(1) Neural Network (MLPRegressor);
(2) Random Forest Regressor;
(3) Linear Regression.

Each model was trained separately for the three pacemaker types using an 80/20 train-test split to ensure performance generalizability.

2.3.1 Neural Network Model

A "Multi-Layer Perceptron (MLP) Regressor" was implemented with three hidden layers (100, 50, and 25 neurons). The ReLU activation function was used, and training was performed using the Adam optimizer. The model was trained for 2000 iterations with a learning rate of 0.001. To decide the best design for our Neural Network, we tested different setups using a method called grid search along with cross-validation, which helps us find what works best while avoiding overfitting. After trying different options, we chose a model with three hidden layers that have 100, 50, and 25 neurons. This "pyramid" shape where each layer has fewer neurons than the one before helps the model learn important patterns step by step while reducing the chances of learning noise from the data. We picked this structure because it gave the best results in terms of accuracy and how well the model matched the real data during testing.

The number of iterations (2000) was chosen based on convergence behavior observed during training. We found that the model typically converged well before 2000 iterations, with minimal improvement beyond that point. Alternative configurations (*e.g.*, 2 layers, different neuron counts, 1000 or 3000 iterations) were evaluated but led to either poorer predictive performance or signs of overfitting.

2.3.2 Random Forest Model

The "Random Forest Regressor" was trained with 150 decision trees. This ensemble method improved accuracy by reducing overfitting and capturing nonlinear relationships in the data. We tested different tree counts (50, 100, 200), and found that 150 trees offered stable, accurate predictions without unnecessary computation. Fewer trees led to less reliable results, while more trees added training time with little benefit. This setup proved most effective for predicting pacemaker battery life across all device types.

2.3.3 Linear Regression Model

A "Linear Regression" model was used as a baseline to evaluate the predictive power of the dataset under the assumption of linear relationships.

2.4 Model Evaluation

Each trained model was evaluated using the following metrics:
(1) Mean Absolute Error (MAE);
(2) Mean Squared Error (MSE);
(3) R-Squared Score (R^2).

Visualization techniques such as residual plots, and normal probability plots, Monte Carlo plots were employed to analyze model performance.

2.5 Monte Carlo Simulation

A "Monte Carlo Simulation" was performed to assess model robustness. Predictions were generated with slight variations in input data introducing random noise into selected features, simulating real-world uncertainty. Each model underwent 1000 iterations, and the distribution of predicted battery life values was analyzed using:
(1) Histogram plots;
(2) Confidence intervals (95%);
(3) Cumulative Distribution Function (CDF) plots.

Hence, this section outlines the methodology used to predict pacemaker battery life for different pacemaker types. Data preprocessing, feature selection, model training, and evaluation methods were described, along with the use of Monte Carlo simulations for reliability analysis.

3 Results

This section presents the results obtained from the predictive models developed for estimating the remaining battery life of single-chamber, dual-chamber, and three-chamber pacemakers. The analysis includes Normal Probability Plots (Q-Q plots) to assess normality, Residual Analysis to evaluate model fit, Monte Carlo Simulation results for uncertainty estimation, and Model Performance Evaluation to compare prediction accuracy. The results for each pacemaker type are presented sequentially.

3.1 Analysis of Each model in different types of Pacemakers

3.1.1 Normal Probability plot

These plots are normal probability plots (Q-Q plots) for the predictions of three different models: Neural Network, Random Forest, and Linear Regression. The Q-Q plot compares the ordered predictions of each model to a theoretical normal distribution, with the red line representing the expected normal distribution. (Figure 4, 5 and 6)

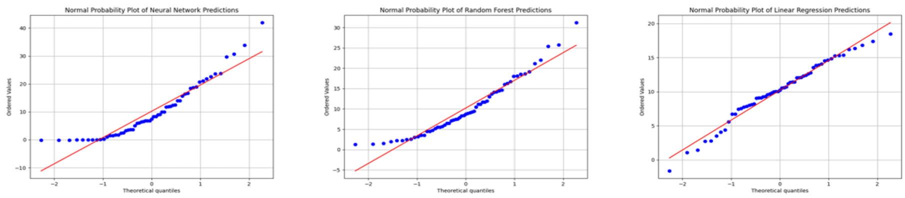

Figure 4 Normal Probability curve (NN, RF, LR) (single-chamber pacemaker)

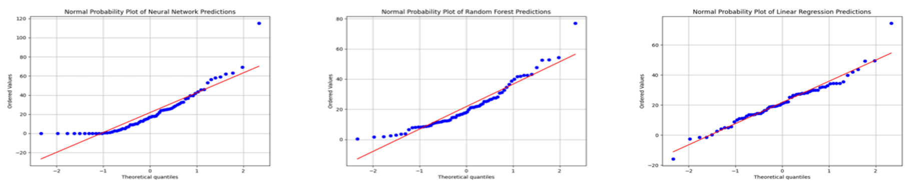

Figure 5 Normal Probability plot (NN, RF, LR) (dual-chamber pacemaker)

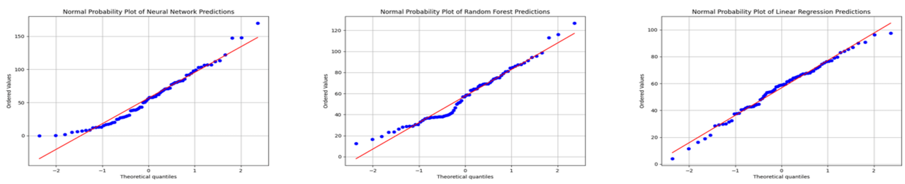

Figure 6 Normal Probability plot (NN, RF, LR) (Triple-chamber pacemaker)

Each plot represents how closely the predictions of a model follow a normal distribution. If the points lie on the red line, then the data closely follows a Gaussian (normal) distribution [35]. Deviations from this line indicate non-normality or heteroscedasticity.

(1) X-axis (Theoretical Quantiles): These are the expected quantiles assuming a normal distribution.

(2) Y-axis (Ordered Values): These are the sorted residuals (prediction errors) from the model.

(3) Red Line: The ideal reference line: if the residuals are normally distributed, the points will lie along this line. Deviations from the red line indicate departures from normality, such as skewness or heavy tails.

Table 1 summarizes the key observations from quantile-quantile (Q-Q) analyses, highlighting the strengths and limitations of each model in terms of statistical assumptions and predictive consistency.

Table 1 Normal Probability plot comparison for different pacemaker types and machine learning models

Pacemaker Type/Model	Neural Network	Random Forest	Linear Regression
Single Chamber	Central residuals ˜ normal; tails show significant deviation. Right skew indicates high variances at high values [36]. Heteroscedasticity present due to extreme prediction errors [37].	Central residuals nearly normal; upper tail shows positive skew. RF overfits extremes. Stepwise nature of trees leads to non-uniform variance in residuals [38].	Residuals closely follow normal distribution. Minor tail deviations. Assumptions of homoscedasticity and normality hold [39,40]. Best statistical reliability.
Dual Chamber	Residuals mostly normal centrally; upper tail has outliers. Positive skew observed. Indicates sensitivity to large values and possible overfitting.	Similar to NN, but fewer extreme outliers. Central alignment better. Upper tail deviation reflects variance instability at high values.	Strongest alignment with normality across all quantiles. Very low residual variance. Linear model generalizes well but may underfit complex patterns.
Triple Chamber	Moderate normality. Long right tail; slight lower curvature. Positive skew. Flexibility leads to extrapolation and variance at high predictions.	Central quantiles align well; tails deviate forming "S-shape." Leptokurtic distribution (heavy tails). Aggregation smooths central predictions but tails still deviate.	Excellent fit to red line. Minor tail deviations. Supports reliable inference. May lack flexibility in modeling nonlinearities in triple chamber signals.

3.1.2 Actual *vs.* Predicted Analysis

In the below actual Vs predicted battery life plots, the X-axis represents the actual battery life (in months), while the Y-axis represents the predicted battery life by the model. These plots help visualize model accuracy, bias, and variance by comparing the closeness of predictions to the ideal 45-degree line (where predicted = actual). Patterns such as systematic under- or overestimation, spread around the diagonal, and slope deviations offer insights into each model's generalization capability and robustness across the full range of observed battery life values. ((Figure 7, 8 and 9))

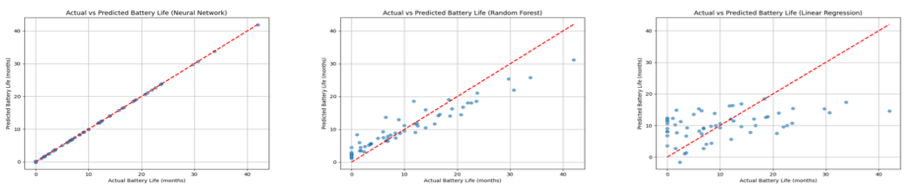

Figure 7 Actual *vs* predicted Analysis (NN, RF, LR) (single-chamber pacemaker)

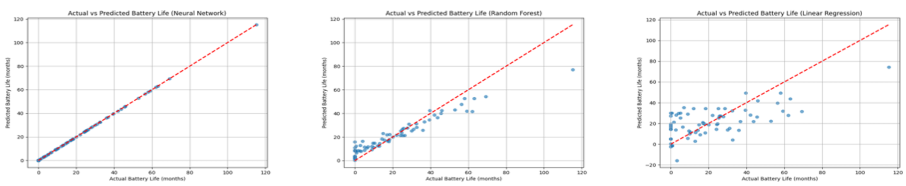

Figure 8 Actual *vs* predicted Analysis (NN, RF, LR) (dual-chamber pacemaker)

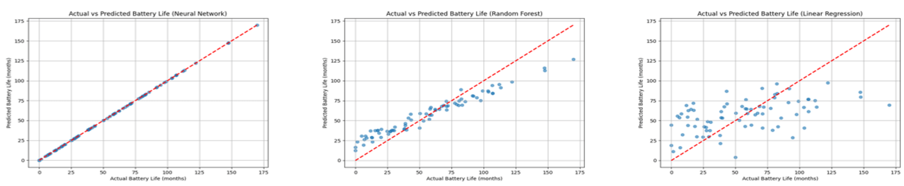

Figure 9 Actual *vs* predicted Analysis (NN, RF, LR) (Triple-chamber pacemaker)

Table 2 summarizes the key findings from these plots for each model–device type combination.

Table 2 Actual *vs.* Predicted plot comparison for different pacemaker types and machine learning models

Pacemaker Type/Model	Neural Network	Random Forest	Linear Regression
Single Chamber	Predictions are highly concentrated along the identity line (y = x), indicating high model fidelity, indicating excellent accuracy. No systematic bias, slope ≈ 1.0, minimal error.	Accurate in mid-range, but overestimates at higher values. Spread increases at high battery life values, possibly due to averaging or data sparsity.	Widest spread, systematically underestimates higher-range target values. Best-fit slope $<$ 1.0, suggests underfitting and inability to capture full data variability.
Dual Chamber	Predictions exhibit near-unity slope alignment ($\hat{y} \approx$ y), confirming model accuracy. Nearly no deviation, very high R^2. Indicates excellent accuracy, but may suggest overfitting or data leakage.	Tends to underpredict high values. Performs well in 0–50 months range. Shows prediction compression at higher values.	Predicted values exhibit high dispersion, deviating substantially from actual targets. Underpredicts high and systematically overestimates lower-bound outcomes. Some predictions are negative, showing poor fit and underfitting.
Triple Chamber	Fit closely approximates the line of equality, with minimal prediction error across the domain. Outstanding accuracy across full range (0–175 months), high $R^2 \approx$ 1.0.	Underpredicts values above ~100 months. Performs well in low-to-mid range. Suffers from tree averaging effect, leading to compression.	Marked deviation from the identity function; high bias and low correlation. Predictions cluster around mid-range, showing underfitting and failure to model complex relationships. Low R^2 and high error.

3.1.3 Residual Analysis

To assess model robustness and the validity of regression assumptions, residuals were plotted against predicted battery life values for each pacemaker type. These plots help identify patterns such as heteroscedasticity, nonlinearity, or bias, which may not be captured through overall error metrics alone. A well-behaved residual plot should show residuals randomly scattered around zero, indicating minimal bias and consistent variance. Conversely, structured patterns (*e.g.*, U-shapes or trends) indicate underfitting, model misspecification, or failure to capture nonlinear relationships.

Residuals were calculated as the difference between the actual and predicted battery life values, defined as: Residual = {Actual Value} – {Predicted Value}.

In residual plot, the X-axis represents the predicted battery life, while the Y-axis shows the residuals (difference between actual and predicted values). ((Figure 10, 11 and 12))

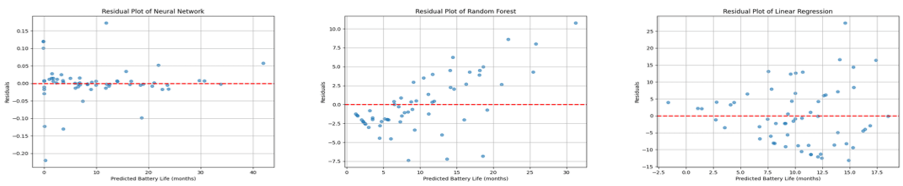

Figure 10 Residual plot NN, RF, LR) (single-chamber pacemaker)

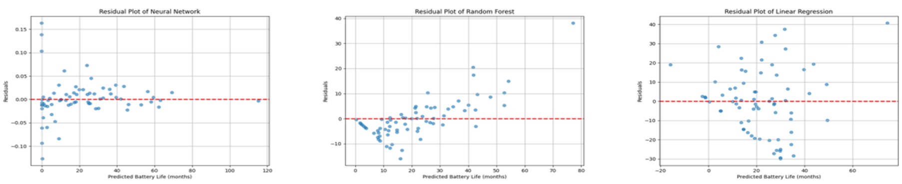

Figure 11 Residual plot (NN, RF, LR) (dual-chamber pacemaker)

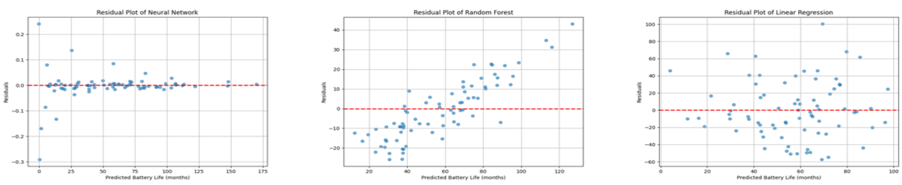

Figure 12 Residual plot (NN, RF, LR) (Triple-chamber pacemaker)

Table 3 summarizes the residual distribution characteristics for each model across single, dual, and triple-chamber pacemakers.

Table 3 Residual plot comparison for different pacemaker types and machine learning models

Pacemaker Type/Model	Neural Network	Random Forest	Linear Regression
Single Chamber	Residuals uniformly scattered around zero; no discernible pattern; low bias and variance; assumption of homoscedasticity satisfied.	Residuals mostly around zero, but increasing variance with higher predicted values indicates heteroscedasticity; possible overfitting in mid-range and poor performance at extremes.	Clear U-shaped pattern; indicates failure to model nonlinearities; strong heteroscedasticity and underfitting; violates assumptions of linearity and independent errors.
Dual Chamber	Residuals tightly clustered around zero (-0.1 to 0.15); minor fan shape at low end; low bias and variance; good fit with minor asymmetry in early predictions.	Residuals increase with predicted battery life; upward trend and outliers; heteroscedasticity evident; underprediction at higher values; may require tuning or feature expansion.	Wide residual range (-35 to +40); U-shaped pattern indicates poor fit; underpredicts at extremes and overpredicts mid-values; significant nonlinearity and underfitting; violates key assumptions.
Triple Chamber	Residuals tightly clustered around zero (-0.3 to 0.2); randomly distributed; homoscedastic; very low error and strong model performance; likely scaled residuals.	Clear positive linear trend; residuals not random: underpredicts low and overpredicts high; systematic bias suggests underfitting or missing features; violates regression assumptions.	Wide residual spread (-60 to +100); no consistent pattern but large errors across all ranges; heteroscedasticity and underfitting evident; fails to model nonlinearity and lacks precision.

3.1.4 Monte Carlo Histogram

Monte Carlo simulations assess the variability in predictions by repeatedly sampling from the model's distribution. These histograms display the distribution of predicted battery life (in months) for three models: Neural Network, Random Forest, and Linear Regression [41]. The x-axis represents the predicted battery life, while the y-axis represents frequency (how often a particular prediction occurs). Each plot includes a kernel density estimate (KDE) curve and a red dashed line representing the mean prediction. ((Figure 13, 14 and 15))

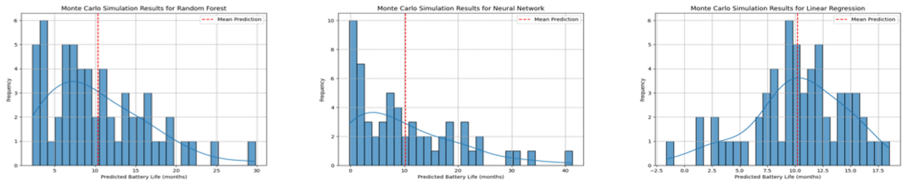

Figure 13 Monte Carlo Histogram (NN, RF, LR) (single-chamber pacemaker)

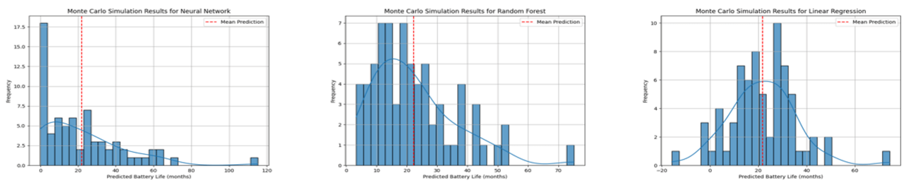

Figure 14 Monte Carlo Histogram (NN, RF, LR) (dual-chamber pacemaker)

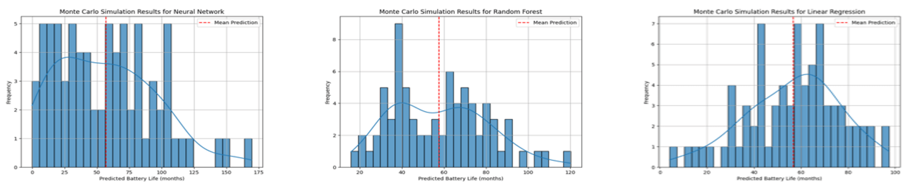

Figure 15 Monte Carlo Histogram (NN, RF, LR) (Triple-chamber pacemaker)

Table 4 analyzes the distribution of predicted battery life values generated by each model, offering insight into model stabilit, bias, variance, and how well predictions conform to expected statistical behavior. It builds on the performance evaluation by focusing on the shape and spread of predictions rather than just their correctness.

3.1.5 Cumulative Distribution Function (CDF)

These plots represent the Cumulative Distribution Functions (CDFs) of the predicted battery life for the Neural Network, Random Forest, and Linear Regression models. The x-axis represents predicted battery life values, and the y-axis represents the cumulative probability. Each CDF plot shows how the predicted values accumulate over the range of predicted battery life, providing insights into the distribution shape, skewness, and concentration of predictions. ((Figure 16, 17 and 18))

Table 4 Monte Carlo Histogram comparison for different pacemaker types and machine learning models

Pacemaker Type/Model	Neural Network	Random Forest	Linear Regression
Single Chamber	Right-skewed distribution with most predictions in the 0–10-month range and some extreme predictions up to 40 months. High variance, skewness >1, kurtosis >3 indicate poor generalization. [42].	Also, right-skewed but less extreme. Predictions mostly within 0–15 months, some up to 30+. Lower variance, smoother KDE curve. More consistent and conservative than NN.	Symmetric, normal-like distribution centered around ˜10 months. Very few outliers. Low variance and kurtosis. High bias and risk of underfitting, but outputs are stable and predictable.
Dual Chamber	Highly right-skewed. Predictions mostly 0–20 months, with some extreme values beyond 100. Mean heavily influenced by outliers. Very high variance possibly dure to large train-test performance gap. Median or trimmed mean more suitable.	Right-skewed but more controlled. Most predictions between 10–40 months. Fewer outliers (up to ˜80 months). Averaging reduces outlier impact. Moderate variance and more robustness.	Bell-shaped, symmetric distribution. Predictions centered around the mean. Adheres to linear assumptions, consistent and reliable, though may underfit complex patterns.
Triple Chamber	Wide, right-skewed, and multi-modal distribution with values up to ˜170 months. High variance, and presence of outliers. Poor generalization and robustness.	Narrower distribution than NN. Predictions mostly 20–120 months with clusters around 40, 60, and 80. Multimodal but smoother. Captures nonlinearity with better control.	Smooth, symmetric, normal-like distribution from 20–100 months. Mean ≈ 60 months. Well-behaved predictions adhering to statistical assumptions. Consistent but may miss nonlinear complexities.

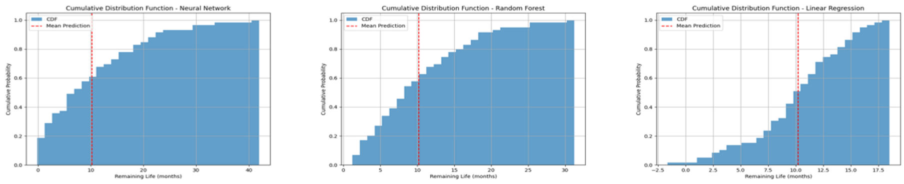

Figure 16 Cumulative Distribution Function (NN, RF, LR) (single chamber pacemaker)

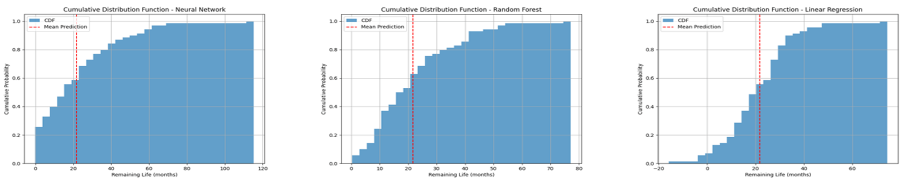

Figure 17 Cumulative Distribution plot (NN, RF, LR) (dual chamber pacemaker)

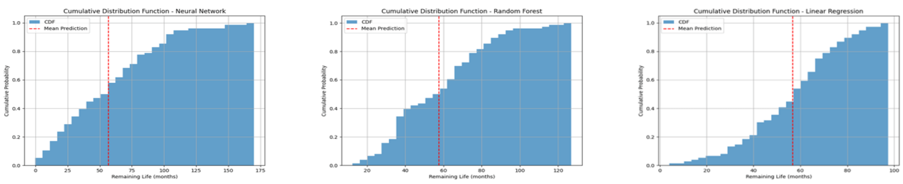

Figure 18 Cumulative Distribution Function (NN, RF, LR) (Triple chamber pacemaker)

Table 5 presents a Cumulative Distribution Function (CDF) analysis of the predicted battery life values for each model and pacemaker type offering insight into prediction concentration, spread, central tendency, and outlier sensitivity.

3.1.6 Confidence Interval 95%

Confidence intervals provide an estimate of the range within which predictions are expected to fall 95% of the time. A 95% confidence interval was chosen because it's a common and trusted standard in both medical research and prediction models. It provides a good balance: it is reliable enough to give confidence in the results, but not so strict that the estimates become too broad to be useful.

When predicting how long a pacemaker battery will last, using a 95% confidence interval means we're pretty sure the true value falls within the range we give 95 times out of 100. That's important when planning for battery replacements, where safety is a top concern. Choosing 99% would be even more cautious, but it would also make the range wider and less precise. On the other hand, 90% would give a narrower range but less certainty. So, 95% is a smart middle ground that fits well with how medical devices are usually analyzed and decisions are made.

The X-axis denotes the sample index, representing individual test data points, while the Y-

Table 5 Cumulative Distribution Function plot comparison for different pacemaker types and machine learning models

Pacemaker Type/Model	Neural Network	Random Forest	Linear Regression
Single Chamber	CDF grows quickly, ˜75% predictions < 25 months; mean ≈ 10 months. Moderate skew with controlled spread and low standard deviation. Suggests a relatively stable model with lower outlier sensitivity.	Most balanced CDF, mean ≈ 10 months at 50th percentile. Linear rise suggests normal-like distribution. Low skew, low variance. High statistical soundness and consistent predictions.	Right-skewed CDF. ˜80% of predictions < 14 months. Mean ≈ 10 months. Wide spread and high variance. Long tail suggests sensitivity to outliers and low precision.
Dual Chamber	Smooth CDF from 0 to 110+ months. Mean ≈ 25 months. Slight right skew, moderate variance. Good nonlinear fit with compressed extreme values. Predictive spread is controlled.	Evenly rising CDF from 0 to <80 months. Mean ≈ 20 months at the center. Symmetrical, tight distribution. Low variance. Predictable and reliable.	CDF from -20 to 60+ months. Mean ≈ 20 months. Moderate right skew with some extreme predictions. High spread, potential bias, and inconsistency in predictions.
Triple Chamber	Gradual rise from ˜10 to 160+ months. Mean ≈ 65 months but skewed right. Wide dispersion reflects uncertainty and large predictive range. Sensitive to high outliers.	Range: 20–120 months. Steeper CDF slope around 60–70 months. Mean ≈ median. Narrow predictive band. Symmetrical, low variance. High confidence model.	S-shaped CDF from 0 to 100 months. Mean ≈ 60 months at 50th percentile. Moderately spread. More centered than other types but still limited in nonlinear capacity.

axis indicates the predicted battery life in months. The blue line represents the mean prediction, and the shaded region corresponds to the 95% confidence interval. ((Figure 19, 20 and 21))

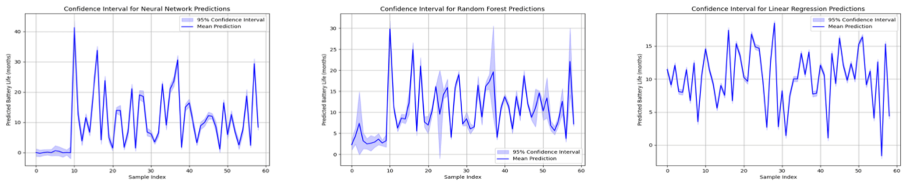

Figure 19 Confidence interval (95%) (NN, RF, LR) (single-chamber pacemaker)

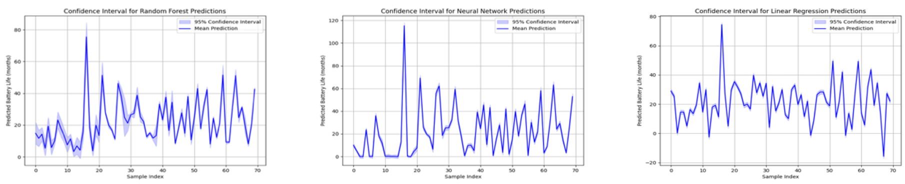

Figure 20 Confidence Interval (NN, RF, LR) (dual-chamber pacemaker)

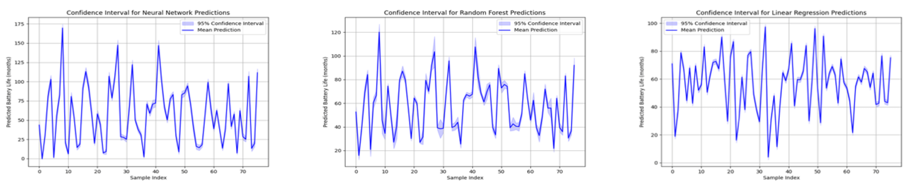

Figure 21 Confidence Interval (95%) (NN, RF, LR) (Triple-chamber pacemaker)

Table 6 summarizes CI behavior across models, highlighting how each algorithm responds to uncertainty in the data, especially under varying degrees of complexity and non-linearity.

3.1.7 Model Evaluation

The performance of three machine learning models: Neural Network, Random Forest, and Linear Regression was evaluated for predicting battery life across single, dual, and triple-chamber pacemakers. Key evaluation metrics include MAE, MSE, and R^2, as summarized in Table 7.

The Neural Network model shows near-perfect performance (MAE: 0.03, MSE: 0.00, R^2: 1.00), indicating excellent prediction accuracy but show minimal bias yet wider variance at extremes, reflecting sensitivity to sparse long duration data. The Random Forest model performs well (MAE: 2.87, MSE: 13.38, R^2: 0.86), though it underestimates high battery life values due to its tree-based structure. In contrast, the Linear Regression model performs poorly (MAE: 6.93, MSE: 74.56, R^2: 0.20), failing to capture the data's complexity and exhibiting high bias and underfitting. (Table 8)

Table 6 Confidence Interval (95%) plot comparison for different pacemaker types and machine learning models

	Neural Network	Random Forest	Linear Regression
Single Chamber	The CI plot exhibits significant variability, with notably wide intervals at indices 10, 25, and 45. This reflects high model uncertainty and instability in predictions. Potential causes include insufficient regularization, or inadequate uncertainty modeling. Uncertainty arises from epistemic (limited data) and aleatoric (due to inconsistencies in interrogation report measurements) factors [43].	The CI plot shows adaptive uncertainty, expanding in high-variance regions. The prediction line fluctuates more than in linear models, reflecting better pattern recognition. However, sharp peaks and dips imply sensitivity to noise. While the model captures complex relationships, its flexibility also introduces localized overconfidence. Ensemble learning reduces but doesn't eliminate variance.	CI plot shows narrow but high-confidence intervals, even when predictions fluctuate significantly. The model fails to capture nonlinearities, leading to underfitting. Narrow intervals indicate overconfidence despite inaccurate predictions. The mean prediction line lacks smoothness, reinforcing the model's inability to generalize. Confidence intervals remain narrow due to the model's analytical estimation under classical assumptions (linearity, homoscedasticity).
Dual Chamber	CI plot shows extremely wide intervals at samples 20 and 25, indicating low prediction confidence. High variance is due to sensitivity to input data variations, and limited training samples. Neural networks lack built-in uncertainty estimation, relying on approximations like MC dropout or bootstrapping. The wide CIs suggest that the model struggles to generalize across test samples.	CI plot is more stable than NN, with moderate and controlled uncertainty. Wider intervals appear at specific samples but without extreme spikes. Ensemble averaging helps reduce prediction variance. The CI width reflects tree disagreement per sample.	CI plot shows the narrowest and most stable intervals. Even near sample extremes, confidence intervals remain tightly bound. This high-confidence behavior results from linear regression's analytical variance estimation under assumptions of residual normality and homoscedasticity. While the model may underfit nonlinear trends, it remains consistent and statistically robust in confidence interval calculation.
Triple Chamber	CI plot shows very high variance and erratic predictions (ranging from near 0 to 170 months). CI bands are often hidden due to overlap with the prediction line. This reflects poor uncertainty estimation and extreme variability. Likely causes include small dataset size, and high epistemic and aleatoric uncertainty. Neural networks lack inherent mechanisms for reliable CI estimation.	CI plot shows moderate variability in CI widths. Predictions are more consistent (20–120 months) than NN. Some indices (10, 25, 40) show wider intervals, indicating localized uncertainty. The model performs ensemble averaging to mitigate variance. Random Forest handles outliers better and is more reliable than NN in estimating uncertainty.	CI plot shows narrow, consistent CIs across all samples (20–90 months). Predictions are smoother and more stable than the other models. High-confidence intervals arise from satisfying classical regression assumptions. While it may not model complex patterns effectively, the linear regression provides trustworthy and analytically sound CIs, making it the most statistically consistent in terms of interval reliability.

Table 7 Model Evaluation (single-chamber Pacemaker)

Model	MAE	MSE	R^2
Neural Network	0.03	0.00	1.00
Random Forest	2.87	13.38	0.86
Linear Regression	6.93	74.56	0.20

Table 8 Model Evaluation (dual-chamber Pacemaker)

Model	MAE	MSE	R^2
Neural Network	0.03	0.00	1.00
Random Forest	5.54	64.77	0.86
Linear Regression	12.99	281.49	0.41

The Neural Network model shows excellent accuracy (MAE: 0.03, MSE: 0.00, R^2: 1.00), but its near-perfect fit suggests possible model complexity issues. The Random Forest model performs well (MAE: 5.54, MSE: 64.77, R^2: 0.86), though it systematically underestimates higher battery life values due to its tendency to flatten extremes. The Linear Regression model performs poorly (MAE: 12.99, MSE: 281.49, R^2: 0.41), indicating it cannot capture the nonlinear patterns in battery life, resulting in high bias and underfitting. (Table 9)

Table 9 Model Evaluation (Triple-chamber Pacemaker)

Model	MAE	MSE	R^2
Neural Network	0.02	0.00	1.00
Random Forest	12.18	221.58	0.85
Linear Regression	25.99	1085.43	0.27

The Neural Network model shows near-perfect accuracy (MAE: 0.02, MSE: 0.00, R^2: 1.00), suggesting excellent pattern capture but raising concerns about high variance in predicting the extreme values. The Random Forest model demonstrated robust predictive capability (MAE: 12.18, MSE: 221.58, R^2: 0.85), though it struggles with extreme values, leading to moderate errors. Linear Regression exhibited limited predictive fidelity (MAE: 25.99, MSE: 1085.43, R^2: 0.27), failing to capture nonlinear patterns and producing highly inaccurate predictions.

3.2 Model Equation

Based on the results, the following equations were derived to represent the relationship between the input features and the predicted remaining battery life for each model:

3.2.1 Neural Network Model

A Multi-Layer Perceptron (MLP) does not provide an explicit equation like a linear regression model [44], but it follows this general structure:

$$\hat{y} = f\left(W_3 \cdot f\left(W_2 \cdot f\left(W_1 \cdot X + b_1\right) + b_2\right) + b_3\right) \tag{1}$$

Where:
(1) X is the input feature vector.
(2) W_1, W_2, W_3 are weight matrices for each layer (100 → 50 → 25).
(3) b_1, b_2, b_3 are biases for each layer.
(4) f (·) f (·) is the ReLU activation function applied to each hidden layer.
(5) The output (y^) is a single continuous value representing Remaining Life (months).

3.2.2 Random Forest Model

A Random Forest consists of multiple decision trees where each tree predicts a value, and the final output is the average of all tree predictions.

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{f_t(x)} \tag{2}$$

Where:
(1) $\hat{y}$ = predicted value for "Remaining Life= $\frac{1}{T}\sum_{t=1}^{T}$ (months).
(2) T = number of trees in the forest (150 in this case).
(3) f_t(x) = prediction of the t^{th} tree for the input features x.
(4) x = vector of input features (*e.g.*, Lead impedance, Capture Threshold, *etc.*)

Random Forest does not have a single closed-form equation, as it is a collection of multiple tree-based models [45, 46].

3.2.3 Linear Regression Model

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots\cdots\cdots\cdots + b_nx_n \tag{3}$$

Where:
(1) $\hat{y}$ = predicted value for "Remaining Life (months).
(2) b_0 = intercept.
(3) b_1, b_2, ········· b_n= regression coefficients (learned from training).
(4) x_1,x_2, ·········x_n = the input feature (*e.g.*, Lead impedance, Capture Threshold, *etc.*)
(5) n = number of features.

4 Discussion

This study comprehensively evaluated three machine learning models**:** Neural Network (NN)**,** Random Forest (RF)**,** and Linear Regression (LR) for their ability to predict pacemaker battery life across single, dual, and triple-chamber devices using interrogation report features. The comparison incorporated statistical metrics (MAE, MSE, R^2) and interpretive visualizations including Actual *vs.* Predicted plots, Residual plots, Normal Probability (Q-Q) plots, Monte Carlo simulations, and Confidence Intervals (CIs).

4.1 Model Performance Across Pacemaker Types

Across all pacemaker types, the Neural Network (NN) model consistently achieved near-perfect R^2 scores (~1.00) (Table 1 to 3), with minimal MAE and MSE, particularly in single-chamber (MAE = 0.03) and dual-chamber (MAE = 0.03) predictions. These results are visually supported by the tight clustering along the diagonal in the Actual *vs.* Predicted plots (Figure 7-9), especially in the triple-chamber case (Figure 9), where NN achieved near-flawless alignment even in the upper tail of battery life predictions.

However, this remarkable accuracy comes with less reliability in extrapolation for battery lifespans beyond the training distribution, particularly evident in the Residual plots (Figure 10-12) and Monte Carlo Histograms (Figure 13-15). In the triple-chamber dataset, NN's residuals, while centralized, show higher variance and heavier right tails. The Monte Carlo simulations reflect this with broad, multi-modal distributions (Figure 15), and the 95% Confidence Interval plots (Figure 19-21) expose erratic intervals, particularly in samples beyond 100 months, which suggests diminished robustness at extrapolative edges.

In contrast, the Random Forest (RF) model maintained high accuracy with greater stability, achieving $R^2 \approx 0.85$ across all device types. Although not as precise as NN, RF predictions displayed better variance control, as seen in smoother KDE curves (Figure 13-15) and narrower CIs in complex triple-chamber predictions (Figure 21). However, RF consistently systematically under-estimated battery life at the higher end, a trend observable in the flattening of predictions above 100 months (Figure 8-9), and confirmed by the positive skew in residuals and the "S-shaped" tail patterns in Q-Q plots (Figure 4-6).

Linear Regression (LR) underperformed across all pacemaker types. It displayed low R^2 values (0.20–0.41)**,** high MAE, and significant residual spread. In the Actual *vs.* Predicted plots (Figure 7-9), LR predictions cluster around the mean, failing to capture variability. Its Residual plots reveal U-shaped patterns and heteroscedasticity, especially in dual and triple chambers (Figure 11-21), indicating violations of model assumptions. Nonetheless, LR produced the narrowest and most stable CIs (Figure 21), a reflection of its statistical simplicity, though at the cost of accuracy and clinical utility.

4.2 Residual Behavior and Distributional Patterns

The Normal Probability plots (Figure 4-6) reinforced model behaviors. While NN and RF deviated in tails, LR residuals aligned most closely with the theoretical normal, supporting its statistical validity but not predictive value. In triple-chamber pacemakers, NN and RF residuals displayed heavier tails, indicating challenges in modeling extreme values likely due to low sample size (n=7) and complex energy consumption patterns.

The Monte Carlo histograms and CDFs (Figure 13-18) further illuminated model characteristics. NN distributions were right-skewed with high kurtosis, especially in the triple-chamber set, implying overconfidence with occasional large deviations. RF distributions were smoother, with multi-modal behavior in complex cases, reflecting robust generalization with cautious prediction. LR distributions, in contrast, were symmetrical but narrow, reflecting a failure to capture nonlinear dynamics, as seen in the CDF's gradual slope and clustering around mean predictions.

4.3 Interpretation of Confidence Intervals and Predictive

Confidence intervals (Figure 19-21) revealed critical differences in uncertainty estimation. NN models, particularly in the triple-chamber set, showed wide and erratic intervals, which aligned with the high residual variance and Monte Carlo spread: signaling poor generalization at data extremes. RF models, in comparison, presented adaptive and moderate-width intervals, suggesting more stable uncertainty estimation across all pacemaker types. LR's intervals, while tight, often misrepresented true uncertainty, especially in underfitted scenarios, as it failed to reflect complex input–output variability.

4.4 Pacemaker Type-Specific Observations

Single-Chamber Pacemakers: All models performed best on this set, owing to lower feature complexity and more stable energy profiles. NN achieved perfect predictions (Figure 7), with minimal residuals and compact CIs. RF slightly systematically under-estimated high-end values, while LR consistently showed systematic bias and underfitting.

Dual-Chamber Pacemakers: NN retained top performance but showed sensitivity to noise, with slightly wider confidence intervals. RF managed better residual dispersion control than in single-chamber cases, though it suffered from underestimating long-duration batteries. LR's predictions were scattered, with negative values in some cases (Figure 8), making it unsuitable for clinical use.

Triple-Chamber Pacemakers: This was the most challenging group. NN handled complexity best, yet variability increased, with wide residuals, non-normality, and broad CIs (Figure 21). RF emerged as a more reliable alternative, balancing accuracy and generalization. LR failed entirely to capture nonlinear load behavior and had the highest error rates (MAE = 25.99, MSE = 1085.43).

4.5 Summary Table of Comparative Model Performance

Table 10 summarizes these findings, highlighting the strengths and limitations of each model in relation to pacemaker complexity.

Table 10 Comparative Model Performance Summary

Model	Pacemaker Type	R^2	Residual Behavior	CI Stability	Overfitting Risk
NN	Single	1.00	Tight, slight skew	Wide	High
RF	Single	0.86	Slight variance	Moderate	Moderate
LR	Single	0.20	U-shaped	Narrow	Low
NN	Dual	1.00	Tight, minor bias	Erratic	Moderate-High
RF	Dual	0.86	Increasing error	Controlled	Moderate
LR	Dual	0.41	Wide, biased	Stable	Low
NN	Triple	1.00	High variance	Unstable	High
RF	Triple	0.85	Smooth, skewed	Adaptive	Moderate
LR	Triple	0.27	Noisy, underfit	Consistent	Low

4.6 Limitations of Each Model

4.6.1 Neural Network

Limitation: While NNs achieved near-perfect predictive performance ($R^2 \approx 1.00$), they exhibited signs of overfitting, particularly evident in extreme cases with higher variance and wide confidence intervals. The model's sensitivity to outliers and possible instability in extrapolation also pose concerns.

Uncertainty: Monte Carlo histograms and confidence intervals revealed that NNs produce predictions with greater variance in some cases, indicating lower reliability for data points outside the training distribution.

Interpretability: Due to their complex structure, NNs are often seen as "black box" models, making it challenging to trace how specific input features influence the output.

4.6.2 Random Forest

Limitation: RF models tended to underestimate extreme battery life values. Their tendency to flatten predictions particularly at the high and low ends of the spectrum limits their ability to generalize in extreme cases.

Residual Trends: The residual plots showed heteroscedasticity and a consistent underestimation for longer battery life values, highlighting limited sensitivity to nonlinearities beyond the central data range.

Prediction Smoothness: Due to the piecewise nature of decision trees, RFs may produce discontinuous predictions, which can be problematic for clinical interpretations requiring smooth estimations.

4.6.3 Linear Regression

Limitation: LR consistently underperformed, especially in capturing the complex and nonlinear interactions inherent in pacemaker battery dynamics. This resulted in low R^2 values (0.20–0.41), substantial residual dispersions, and poor alignment in prediction plots.

Assumptions Violated: The model's assumption of linearity was not valid for this application. Residual plots revealed U-shaped patterns and heteroscedasticity, indicating poor fit and significant underfitting.

Predictive Utility: While highly interpretable and stable, LR lacked the flexibility to accommodate the diversity of input patterns across different pacemaker types.

4.7 Conditions for Neural Network Model Success

The Neural Network Model worked very well, but only under certain conditions:

Lots of Good Data: It needed detailed and high-quality information like lead impedance, pacing thresholds, and patient activity to understand the complex behavior of pacemaker batteries.

Clean and Well-Prepared Data: The data had to be carefully prepared. That means filling in missing values, scaling numbers properly, and making sure everything was in the right format. Neural networks are very sensitive to messy or unprocessed data.

Enough Data That Covers All Cases: The model performed best when it was trained on a wide variety of battery life examples not just average cases, but also low and high extremes. If the training data didn't include these, the model became less accurate and more uncertain.

Methods to Control Overfitting: Neural networks can sometimes memorize the training data too well, which hurts their performance on new data. To avoid this, techniques like Monte Carlo simulations and model averaging were helpful. These methods also gave a better idea of how confident the model was in its predictions.

4.8 Clinical Importance

Our model's predictions are very close to what Medtronic already estimates for pacemaker battery life, and this has several important benefits in real-life healthcare:

Proves It Works Independently: It shows that we can accurately estimate battery life using our own machine learning model, without relying on the device maker's internal tools. This adds an extra layer of confidence by confirming their results from the outside.

Helps Compare Different Brands: This model can be used as a fair, vendor-neutral way to compare battery life predictions across pacemakers from different companies. That makes it easier for doctors to make informed choices, no matter which brand they're using.

Supports Remote Monitoring: The model can also help spot unusual battery drain early, especially in patients being monitored remotely. This is helpful because early warning signs can be missed without regular in-person checks.

5 Conclusion and Future work

This study presents a data-driven framework for predicting pacemaker battery life using machine learning models trained on interrogation report data. By applying Neural Networks, Random Forests, and Linear Regression to a structured set of electrical and functional device parameters, we demonstrate that nonlinear models, particularly Neural Networks, consistently outperform simpler approaches in forecasting remaining battery life across single-, dual-, and triple-chamber devices.

The findings confirm that model accuracy improves with increasing complexity, provided the input features are representative of real-world pacing behavior and battery load. Monte Carlo simulations and confidence interval analyses further reinforce the robustness and potential clinical utility of the proposed models. Importantly, while this study utilized data from a single manufacturer, the modelling techniques are applicable to any pacemaker system utilizing lithium-based batteries with similar interrogation parameters.

Looking ahead, future work should focus on:
(1) External validation with multi-manufacturer datasets to test generalizability across device platforms.
(2) Real-time model integration with remote monitoring systems to support dynamic prediction updates.
(3) Inclusion of patient-specific variables (*e.g.*, age, comorbidities, activity level) to enhance personalization.
(4) Development of clinical decision-support tools that translate predicted battery life into actionable scheduling recommendations.

By refining these predictive tools and validating them in broader contexts, we move closer to enabling proactive pacemaker management, reducing unnecessary replacements, and ultimately improving patient outcomes and healthcare efficiency.

Conflicts of interest

The authors declare that they have no conflict of interest.

References

[1] Di Cesare M, Perel P, Taylor S, et al. The Heart of the World. Global Heart. 2024, 19(1).
https://doi.org/10.5334/gh.1288

[2] Murray JA. Pacemaker Therapy for Cardiac Arrhythmias. Archives of Internal Medicine. 1969, 123(4): 355.
https://doi.org/10.1001/archinte.1969.00300140001001

[3] Bularga A. Modern pacemaker technology: enhancing heart rhythm management. Curr Trend Cardiol. 2024, 8(3): 258.
https://doi.org/10.35841/aacc-8.3.258

[4] Burkhard S, Van Eif V, Garric L, et al. On the Evolution of the Cardiac Pacemaker. Journal of Cardiovascular Development and Disease. 2017, 4(2): 4.
https://doi.org/10.3390/jcdd4020004

[5] Eagle KA, Mulley AG, Singer DE, et al. Single-Chamber and Dual-Chamber Cardiac Pacemakers. Annals of Internal Medicine. 1986, 105(2): 264-271.
https://doi.org/10.7326/0003-4819-105-2-264

[6] Gould PA, Mariani JA, Kaye DM. Biventricular pacing in heart failure: a review. Expert Review of Cardiovascular Therapy. 2006, 4(1): 97-109.
https://doi.org/10.1586/14779072.4.1.97

[7] Dretzke J, Toff WD, Lip GY, et al. Dual chamber versus single chamber ventricular pacemakers for sick sinus syndrome and atrioventricular block. Cochrane Database of Systematic Reviews. 2004, 2021(4).
https://doi.org/10.1002/14651858.cd003710.pub2

[8] Pedersen A. Medtronic Gives Heart Docs Something to 'MARVEL' At. 2016.
https://www.mddionline.com

[9] Connolly SJ, Kerr C, Gent M, et al. Dual-Chamber Versus Ventricular Pacing: Critical Appraisal of Current Data. Circulation. 1996, 94(3): 578-583.
https://doi.org/10.1161/01.CIR.94.3.578

[10] Evera MRI TM SureScan ® ICD.
http://spo.escardio.org

[11] Jaffe LM, Morin DP. Cardiac resynchronization therapy: history, present status, and future directions. Ochsner journal. 2014, 14(4): 596-607.
https://www.ochsnerjournal.org

[12] Medical Product Outsourcing. CE Mark for Medtronic's Next Generation CRT Pacemakers. February 24, 2017.

[13] Rosenthal LS, Mester S, Rakovec P, et al. Factors influencing pacemaker generator longevity: results from the complete automatic pacing threshold utilization recorded in the Capture trial. Pacing and clinical electrophysiology. 2010, 33(8): 1020-1030.
https://doi.org/10.1111/j.1540-8159.2010.02809.x

[14] Cardiac Rhythm Disease Management Product Performance Report Important Patient Management Information for Physicians.
https://www.CRDMPPR.medtronic.com

[15] Single Chamber Temporary PaceMaker (AAI/VVI) Technical Manual. 1981.

[16] Explant of Implantable Devices at End of Service CRM Technical Services Standard Letter, 2024.
https://www.medtronic.com

[17] 5392 Dual Chamber Temporary External Pacemaker Technical Manual.

[18] Mallela VS, Ilankumaran V, Rao NS. Trends in cardiac pacemaker batteries. Indian Pacing and Electrophysiology Journal. 2004, 4(4): 201.
https://pmc.ncbi.nlm.nih.gov/articles/PMC1502062

[19] Neupane S, Goswami T. Pacemaker Interrogation Reports: Comparing Diagnostics, Lead Impedance, Pacing Thresholds and Battery Performance. Research on Intelligent Manufacturing and Assembly. 2025, 4(1): 144-167.
https://doi.org/10.25082/rima.2025.01.002

[20] Truong VT, Beyerbach D, Mazur W, et al. Machine learning method for predicting pacemaker implantation following transcatheter aortic valve replacement. Pacing and Clinical Electrophysiology. 2021, 44(2): 334-340.
https://doi.org/10.1111/pace.14163

[21] Severson KA, Attia PM, Jin N, et al. Data-driven prediction of battery cycle life before capacity degradation. Nature Energy. 2019, 4(5): 383-391.
https://doi.org/10.1038/s41560-019-0356-8

[22] Liao SH, Chu PH, Hsiao PY. Data mining techniques and applications – A decade review from 2000 to 2011. Expert Systems with Applications. 2012, 39(12): 11303-11311.
https://doi.org/10.1016/j.eswa.2012.02.063

[23] Ellis CR, Dickerman DI, Orton JM, et al. Ampere Hour as a Predictor of Cardiac Resynchronization Defibrillator Pulse Generator Battery Longevity: A Multicenter Study. Pacing and Clinical Electrophysiology. 2016, 39(7): 658-668.
https://doi.org/10.1111/pace.12831

[24] Boriani G, Merino J, Wright DJ, et al. Battery longevity of implantable cardioverter-defibrillators and cardiac resynchronization therapy defibrillators: technical, clinical and economic aspects. An expert review paper from EHRA. EP Europace. 2018, 20(12): 1882-1897.
https://doi.org/10.1093/europace/euy066

[25] Jimoh ST, Al-Juboori SS. Cyber-Securing Medical Devices Using Machine Learning: A Case Study of Pacemaker. Journal of Informatics and Web Engineering. 2024, 3(3): 271-289.
https://doi.org/10.33093/jiwe.2024.3.3.17

[26] Pan H, Feng W, Ge J, et al. Prediction method of battery remaining life based on CEEMDAN-RF-SED-LSTM neural network. International Journal of Information and Communication Technology. 2024, 25(9): 1-21.
https://doi.org/10.1504/ijict.2024.142816

[27] Kamul A, Hananda N, Harito C, et al. Machine Learning Application in Battery Prediction: A Systematic Literature Review and Bibliometric Study. Mursitama TN, Noerlina, Sitepu E, Basaria FT, eds. E3S Web of Conferences. 2023, 388: 01021.
https://doi.org/10.1051/e3sconf/202338801021

[28] Papadopoulos CE, Yeung H. Uncertainty estimation and Monte Carlo simulation method. Flow Measurement and Instrumentation. 2001, 12(4): 291-298.
https://doi.org/10.1016/s0955-5986(01)00015-2

[29] Velikova T, Mileva N, Naseva E. Method "Monte Carlo" in healthcare. World Journal of Methodology. 2024, 14(3).
https://doi.org/10.5662/wjm.v14.i3.93930

[30] Crossley GH, Mason PK, Hansky B, et al. High predicted durability for the novel small-diameter OmniaSecure defibrillation lead. Heart Rhythm. 2025, 22(2): 302-310.
https://doi.org/10.1016/j.hrthm.2024.09.005

[31] Verma N, Getenet S, Dann C, et al. Evaluating an Artificial Intelligence (AI) Model Designed for Education to Identify Its Accuracy: Establishing the Need for Continuous AI Model Updates. Education Sciences. 2025, 15(4): 403.
https://doi.org/10.3390/educsci15040403

[32] Hodson TO. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development. 2022, 15(14): 5481-5487.
https://doi.org/10.5194/gmd-15-5481-2022

[33] Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. 2005, 30: 79-82.
https://doi.org/10.3354/cr030079

[34] Ujwal S. Regressing metrics using advance machine learning. 2024.
https://www.irjet.net

[35] Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis for the analysis of data. Biometrika. 1968, 55(1): 1-17.
https://doi.org/10.1093/biomet/55.1.1

[36] Boylan GL, Cho BR. The Normal Probability Plot as a Tool for Understanding Data: A Shape Analysis from the Perspective of Skewness, Kurtosis, and Variability. Quality and Reliability Engineering International. 2011, 28(3): 249-264.
https://doi.org/10.1002/qre.1241

[37] Larasati A, Hajji AM, Dwiastuti A. The relationship between data skewness and accuracy of Artificial Neural Network predictive model. IOP Conference Series: Materials Science and Engineering. 2019, 523(1): 012070.
https://doi.org/10.1088/1757-899x/523/1/012070

[38] Barreñada L, Dhiman P, Timmerman D, et al. Understanding overfitting in random forest for probability estimation: a visualization and simulation study. Diagnostic and Prognostic Research. 2024, 8(1).
https://doi.org/10.1186/s41512-024-00177-1

[39] Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International. 2010, 107(44): 776.

[40] Solutions S. Statistics solutions: Advancement through clarity. 2018, 5.
http://www.statisticssolutions.com

[41] Zhang Y, Wik T, Bergström J, et al. A machine learning-based framework for online prediction of battery ageing trajectory and lifetime using histogram data. Journal of Power Sources. 2022, 526: 231110.
https://doi.org/10.1016/j.jpowsour.2022.231110

[42] Hair Jr JF, Hult GTM, Ringle CM, et al. Partial least squares structural equation modeling (PLS-SEM) using R: A workbook. Springer Nature, 2021.

[43] Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine Learning. 2021, 110(3): 457-506.
https://doi.org/10.1007/s10994-021-05946-3

[44] Gaudart J, Giusiano B, Huiart L. Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data. Computational Statistics & Data Analysis. 2004, 44(4): 547-570.
https://doi.org/10.1016/s0167-9473(02)00257-8

[45] Reis I, Baron D, Shahaf S. Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets. The Astronomical Journal. 2018, 157(1): 16.
https://doi.org/10.3847/1538-3881/aaf101

[46] Klusowski JM. Sharp analysis of a simple model for random forests, International Conference on Artificial Intelligence and Statistics. PMLR. 2021: 757-765.
http://arxiv.org/abs/1805.02587