Engineering and Applied Science Research
https://www.tci-thaijo.org/index.php/easr/index
Published by the Faculty of Engineering, Khon Kaen University, Thailand

Evaluation of missing value handling methods in machine learning for emergency department mortality prediction

Narawish Kophimai[1, 2], Krisanarach Nitisiri*[2], Pariwat Phugoen[3], Kanchana Sethanan[2] and Kuo-Jui Wu[4]

[1]Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[2]Research Unit on System Modeling for Industry, Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[3]Department of Emergency Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand
[4]International Business School, Hainan University, Haikou, China

Received 3 April 2025
Revised 29 May 2025
Accepted 1 August 2025

Abstract

Missing data remains a significant challenge in emergency medicine, particularly in mortality prediction models. This study investigates five distinct missing value handling methods applied to various machine learning algorithms using a dataset of 331,151 emergency department records from a Thai hospital (2016–2021). The study evaluates complete case analysis, zero imputation, mean imputation, k-Nearest Neighbors (kNN) imputation, and MissForest, combined with logistic regression, decision tree, random forest, Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost). The results indicate that XGBoost with zero imputation delivers the best performance, achieving an accuracy of 0.8659, precision of 0.8726, recall of 0.8659, F1-score of 0.8681, and an AUC ranging from 0.8848 to 0.9947 across eight prediction classes. Furthermore, tree-based models demonstrated greater stability across different missing value handling methods, whereas linear models were more sensitive to imputation techniques. These findings suggest that strategic selection of missing data handling approaches can significantly enhance the reliability of mortality predictions in emergency care settings.

Keywords: Mortality prediction, Missing value, Machine learning in emergency medicine, Emergency department, Emergency triage

1. Introduction

The over-triaging of patients in emergency departments (EDs) has emerged as a major issue for resource constraints and operational inefficiencies [1]. This issue contributes to broader challenges of ED overcrowding, which can result in inadequate patient outcomes [2]. When combined with escalating demands for emergency care, over-triaging creates significant bottlenecks, straining the already limited available resources and leading to their suboptimal utilization [3].

To solve the over-triaging issue, machine learning (ML) has become a pervasive technology in contemporary society, finding application across a diverse array of disciplines. The fundamental principle underlying ML is the utilization of historical data to enable autonomous learning and the prediction of outcomes with unseen data. A significant advantage of ML is its capacity to mitigate biases arising from human error [4]. Numerous applications have been developed within the medical field, including prediction of heart failure and cancer or tumor detection through image classification [5]. Recent advancements in the field of emergency medicine have yielded a wide range of treatment options and enhanced patient outcomes, particularly in cases of urgent health concerns. ML algorithms have been employed in emergency medicine to predict mortality outcomes and admissions [6]. However, only a limited number of studies have been conducted to predict the multiple outcomes of emergency triage like death in emergency room (ER), death within 24 hours, admission to the intensive care unit (ICU), and in-hospitality admission.

Missing data is a common issue in real-world clinical datasets. It can be caused by human error, differences in documentation practices, equipment failure or patient conditions that prevent data collection. The presence of missing data within a specific dataset can be so pervasive that the deletion of the record containing the missing value becomes unfeasible [7]. Addressing missing data can be achieved through various techniques, including deletion, ignoring, estimation of statistical values, and imputation [8]. In high-pressure environments such as EDs, incorrect or biased predictions resulting from unaddressed missing values can result in poor clinical decisions.

Despite the significance of this problem, existing studies often overlook the impact of different missing value handling strategies on model performance—particularly in multi-class prediction settings where the stakes vary across outcomes like ICU admission or short-term mortality. Moreover, many studies focus on accuracy, neglecting other critical evaluation metrics such as recall and precision that are vital in clinical contexts.

To address these limitations, this study investigates the impact of five missing value handling methods on the performance of five machine learning models in predicting eight different emergency department mortality and admission outcomes. The objective is to

*Corresponding author.
Email address: krisni@kku.ac.th
doi: 10.14456/easr.2025.48

identify effective imputation techniques that can enhance model robustness and predictive performance in the presence of incomplete data. The findings aim to guide future model development for real-world emergency care environments where data quality issues are inevitable.

The following sections of this research will build upon the previously reviewed literature concerning mortality prediction with machine learning and methods for handling missing data on the dataset. Section 3 will describe methodology of the research including dataset description, data preparation, imputation methods, ML models and evaluation metrics. Section 4 will present the results of the computational analysis. Section 5 will provide a discussion of the research's results. Finally, section 6 will offer a summary of the research's findings and their future implications.

2. Literature review

2.1 Machine learning in emergency medicine

A variety of features were employed in the prediction of mortality, including age, sex, vital signs (e.g., body temperature, heart rate, blood pressure, and oxygen saturation), and transportation method (e.g., self-entrance or emergency van) [9-11]. Studies have used machine learning to predict the mortality and triaging scores.

Several studies have successfully demonstrated the effectiveness of ML in predicting binary clinical outcomes. For example, logistic regression and random forest have been applied to predict 7-day and 30-day mortality, achieving AUCs above 0.90 in training sets [10]. Similarly, deep learning and ensemble tree models have outperformed conventional triage systems in estimating Emergency Severity Index (ESI) scores and predicting ICU admissions [11].

However, many of these models are limited to binary classification and do not account for the range of clinically relevant outcomes. A study highlighted the need for multi-class prediction models in ED settings to better reflect the complexity of patient trajectories [12]. Moreover, most of the reviewed studies assume complete data availability, despite the reality that EHRs are often incomplete.

Thus, while prior research supports the potential of ML in emergency medicine, there remains a gap in the development of models that can handle multiple outcomes under real-world data conditions, particularly when dealing with missing values.

2.2 Missing value handling methods in Electronic Health Record (EHR) dataset

Based on probability and distribution, the missing data can be divided into three main mechanisms: not missing at random (NMAR), missing at random (MAR), and missing completely at random (MCAR). Missingness that is unrelated to observed or missing data is a characteristic of MCAR. On the other hand, missingness is linked to observed data but not to missing data, according to MAR. Lastly, NMAR indicates that missing values in the data matrix directly affect missing data [13]. Traditional methods for handling missing data, such as complete case analysis or mean imputation, are simple to apply but can lead to biased results or significant information loss when the assumptions of MCAR are violated [14].

In response to these limitations, more advanced imputation methods have emerged. MissForest, which uses random forests to estimate missing values has shown improved performance over statistical methods, especially in high-dimensional medical datasets [15]. A recent study [16] analyzed EHR data from diabetic patients and found that multiple imputation using chained equations (MICE) yielded better prediction accuracy than complete case analysis, highlighting the importance of robust imputation in clinical settings.

Interpretable machine learning has also been explored in the context of missing data. A study demonstrated how explainable boosting machines (EBMs) can help uncover missingness mechanisms and guide imputation strategy selection [17]. Another study used the Framingham Heart Study dataset, compared traditional imputation methods with deep learning techniques such as autoencoders and generative adversarial networks (GANs) [18]. Their findings showed that deep learning models better preserved the underlying data structure, particularly in clinical domains.

Moreover, studies have explored optimization-based approaches. For example, hybrid methods using whale optimization and late acceptance hill climbing have improved imputation for diabetic patient datasets [19]. Similarly, multi-objective particle swarm optimization (MOPSO) has been applied to maximize sensitivity and specificity in cancer datasets with missing data [20].

While these methods show promise, most studies focus on improving a single performance metric—typically accuracy—without considering precision or recall. This limits their clinical relevance, as false negatives in mortality prediction can have severe consequences. Moreover, few works evaluate these techniques across multiple ML models and multiclass outcomes, which is critical in emergency triage where patients can follow diverse trajectories (e.g., death in a day, ICU admission, or discharge).

2.3 Research gap

Despite the increasing integration of machine learning in emergency medicine, several critical gaps remain in the literature—particularly regarding the treatment of missing data in multiclass clinical prediction tasks.

First, most prior research assumes a complete dataset or only superficially addresses missingness using simplistic strategies such as complete case analysis or mean imputation. While these methods are easy to implement, they often introduce bias, reduce statistical power, or fail to reflect real-world data conditions where missing values are pervasive and non-random. Although some advanced imputation methods have been proposed—such as MissForest, kNN, or optimization-based techniques—there is limited comparative evaluation across multiple methods within the same experimental framework. This makes it difficult to determine which techniques are most robust or suitable for high-stakes clinical applications.

Second, much of the existing literature focuses on binary outcomes (e.g., survival vs. death, admit vs. discharge), which oversimplifies the spectrum of patient conditions in emergency departments. Real-world scenarios often involve a variety of possible outcomes, such as death within different time windows (1, 3, 7, or 60 days), ICU admission, or discharge to general wards. These outcomes have different implications for hospital resource allocation and patient management but are rarely captured together in machine learning models.

Third, while some studies report accuracy, fewer consider additional performance metrics such as precision, recall, and AUC, which are essential for evaluating predictive models in clinical environments where false negatives or false positives can lead to serious consequences. Additionally, performance stability across imputation methods and model types is rarely analyzed—yet this is vital in healthcare settings where missing data mechanisms may change over time or vary between institutions.

To address these gaps, this study presents a large-scale, multiclass evaluation of five missing value handling methods—ranging from simple substitutions including complete case analysts, zero imputation and mean imputation to algorithmic imputations including, k-NN imputation and MissForest—across five machine learning models including logistic regression, decision tree, random forest, Light Gradient Boosted Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost). Using real-world emergency department data with eight clinically relevant outcome classes, the analysis considers not only accuracy, but also precision, recall, F1-score, and AUC. This comprehensive approach provides robust insights into which imputation techniques and model combinations offer the most reliable performance under the constraints of real-world clinical data.

3. Methodology

The experiment is coded using Python 3.10.13, which was executed locally on an M3 Apple silicon CPU. To scale and generate models, the Scikit-Learn, LightGBM and XGBoost libraries were utilized. The dataset's imbalanced class was addressed using the Imbalanced-learn library.

3.1 Dataset description

The dataset was retrieved from an ED of a hospital located in north-eastern region of Thailand, comprising a total of 331,151 records during the period from 2016 to 2021. dataset consists of eight classes and 20 features including vital signs, laboratory tests, nurses' triage, procedure complexity evaluation, and patient-transfer types. The characteristics and missingness of each feature are presented in Table 1. The number of eight prediction classes which are death in a day, three days, seven days, 28 days, 60 days, Intensive Care Unit (ICU), admit, and others, are 383, 309, 284, 622, 182, 7,394, 46,019 and 275,958, respectively.

Table 1 Feature characteristics and their missingness

Feature	n = 331,151	Missingness (%)
Age, median	38	0
Sex, mode(n)	Female (185,676)	0
Patient Type, mode(n)	Non Trauma (279,265)	0
Temperature, mean	36.93	9.7140
PR, mean	91.99	4.4242
RR, mean	20.68	4.4309
SBP, mean	130.02	6.2488
DBP, mean	77.51	6.2555
MAP, mean	89.47	0.0012
O_2 Sat., mean	97.89	10.8654
Eye score, mode(n)	4 (277,066)	15.0542
Voice score, mode(n)	5 (275,203)	15.1375
Movement score, mode(n)	6 (276,609)	15.1737
Triage, mode(n)	Level 4 Less Urgency (132,299)	0
Cardiac arrest, mode(n)	No (330,611)	0
EKG, mode(n)	No (267,010)	1.7360
Plain radiography, mode(n)	No (212,187)	1.7361
CT scan, mode(n)	No (316,642)	0
Blood test, mode(n)	No (198,985)	0
EMS	No (315,650)	0

PR = Pulse Rate, RR = Respiratory Rate, SDP = Systolic Blood Pressure, DBP = Diastolic Blood Pressure, O_2 Sat = Oxygen Saturation, EKG = Electrocardiogram, CT scan = Computerized Tomography scan, EMS = Emergency Medical Services

3.2 Data preprocessing

The categorical features in the dataset were encoded using One-hot encoding method. Subsequently, the data matrix was scaled using Min-Max scaler. Then, the missing values were addressed using each missing value handling method. Once the missing values were handled, the dataset was split into training (80%) and testing (20%) subsets. Then, Synthetic Minority Over-sampling Technique (SMOTE) [21] was used only on the training set to address class imbalance and to keep the test set's original distribution. This approach ensures that oversampling does not introduce synthetic data into the test set, preserving its original distribution. The preprocessing pipeline is illustrated in Figure 1.

3.3 Missing value handling method

A variety of missing value handling methods have been developed to address the issue of missing data points in datasets. These methods encompass a range of approaches, from the straightforward, such as complete case analysis, to the more complex, including single imputation methods like zero imputation or mean imputation, [13] and algorithmic imputation techniques like KNN imputation [22] or MissForest [23].

3.3.1 Complete case analysis

The term "complete case analysis" refers to a methodological approach that involves the exclusion of missing values, thereby simplifying the handling of incomplete datasets by eliminating the samples that contain missing elements.

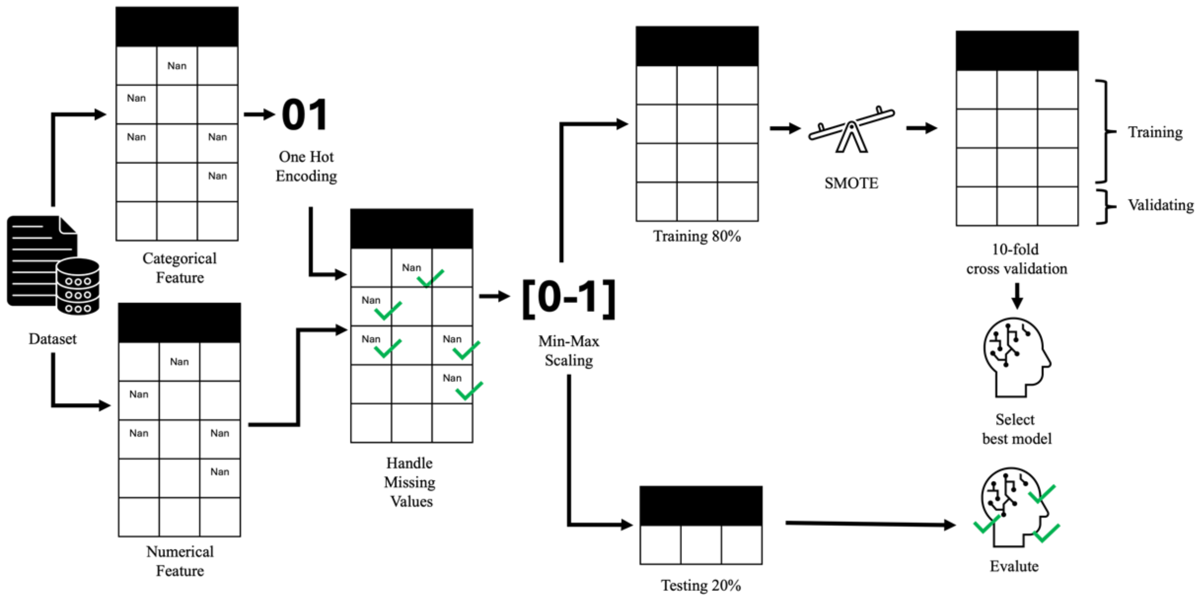

Figure 1 Overall framework of the research

3.3.2 Zero imputation

Zero imputation is a missing value handling that involves replacing any missing values in the dataset with the value of zero.

3.3.3 Mean imputation

Mean imputation is a classical approach in statistical methodology for handling missing data. The underlying principle is straightforward yet powerful: it replaces missing values with the arithmetic mean of the observed values for that variable.

3.3.4 k-Nearest Neighbors (KNN) imputation

The k-nearest neighbors imputer is predicated on the principle that data points sharing similar characteristics will likely possess analogous values. The imputer methodically identifies the k most similar data points (neighbors) that are complete, subsequently utilizing these data points as a reference to estimate the value of missingness [22].

3.3.5 MissForest

MissForest is a non-parametric imputation algorithm that can handle missing values of different types. It employs a random forest model to predict a single missing point in a dataset. The algorithm's use of multiple trees, or bootstrapping, helps to reduce overfitting and improve the accuracy of the predictions [23].

3.4 Machine learning models

Five machine learning algorithms were used in this study, including tree-based models like decision trees, random forests, light gradient boosting machine (LightGBM), and extreme gradient boosting (XGBoost) trees, as well as linear models like logistic regression. These algorithms were implemented with 10-fold cross-validation without replacement to select the optimal combination of ML model and missing value handling method.

3.4.1 Logistic regression

Logistic regression represents a particular type of linear model that has been developed for the purposes of classification. This model involves the calculation of prediction probabilities, which are determined using coefficients in linear equations [24].

3.4.2 Decision tree

A non-parametric machine learning algorithm known as Decision Tree utilizes the advantages of the recursive feature space partitioning principle. This process involves the division of feature spaces into nodes and leaves, a process that continues until the information gain is maximized or the impurity measures are minimized [25].

3.4.3 Random forest

The random forest methodology extends the concept of decision trees through the ensemble methodology, employing the bootstrap aggregating (Bagging) technique combined with a random feature selection process. The final classification prediction is determined by the majority vote of the ensemble members [26].

3.4.4 Light Gradient Boosting Machine (LightGBM)

The LightGBM, an ensemble learning decision tree, employs a light gradient boosting strategy that has been shown to enhance model performance while minimizing memory usage and training time. This approach utilizes a histogram-based algorithm to construct the decision tree, a method that has been demonstrated to optimize operational efficiency. [27].

3.4.5 Extreme Gradient Boosting (XGBoost)

The Extreme Gradient Boosting (XGBoost) tree is a popular machine learning model designed for distributed optimization. It has been utilized in various applications due to its capacity for high-efficiency and adaptability. The XGBoost algorithm employs a gradient-boosting tree architecture, which facilitates parallel computing, resulting in enhanced accuracy and speed for machine learning tasks. [28].

3.5 Evaluation metrics

The evaluation of ML algorithms employed various metrics, including accuracy, weighted precision, weighted recall, and weighted F1 score. Accuracy is defined as the extent to which a model correctly predicts values. Precision, also known as PPV, signifies the performance of a model to accurately predict positive outcomes. Recall, alternatively, is the frequency with which a model identifies true positive instances among all positive instances. The F1 score is a metric that balances precision and recall, offering a comprehensive assessment of the model's performance. In this research, weighted precision, recall and F1-score will be used throughout the analysis. Moreover, Receiver Operating Characteristics (ROC) and Area Under ROC (AUC) curves of each class will be assessed with the selected combination ML model and missing value handling methods.

4. Results

4.1 Performance of missing value handling methods

As shown in Table 2 and visualized in Figure 2 to Figure 5, the results of 10-fold cross-validated metrics tested on five machine learning models including logistic regression, decision tree, random forest, light gradient boosted machine and extreme gradient boosting applied with five missing value handling methods which are complete case analysis, zero imputation, mean imputation, kNN imputation and MissForest are presented. The best-performed combination between ML model and missing value handling method is random forest with kNN imputation with the accuracy of 0.9806, precision of 0.9812, recall of 0.9806, and F1-score of 0.9806.

Moreover, the results of testing dataset show that combining logistic regression model with any missing value handling method outperformed the complete case analysis one by approximately 2%. The best performed pair of random forest and kNN imputation in the training and validation lacks a short amount of performance in all metrics when contrasted with the combination of XGBoost and zero imputation, yielding 0.8659, 0.8726, 0.8659, and 0.8681 in accuracy, precision, recall, and F1-score, respectively.

Table 2 Training and testing results of each ML model and missing value handling method

Model	**Method**	**Training**				**Testing**			
		accuracy	**precision**	**recall**	**f1-score**	**accuracy**	**precision**	**recall**	**f1-score**
Logistic Regression									
	Complete Case Analysis	0.4485	0.4430	0.4485	0.4373	0.7126	0.8688	0.7126	0.7726
	Zero Imputation	**0.4701**	**0.4666**	**0.4701**	**0.4621**	0.7335	**0.8739**	0.7335	**0.7876**
	Mean Imputation	0.4674	0.4620	0.4674	0.4584	**0.7339**	0.8737	**0.7339**	0.7868
	KNN Imputation	0.4664	0.4621	0.4664	0.4583	0.7325	0.8736	0.7325	0.7864
	MissForest	0.4674	0.4620	0.4674	0.4584	**0.7339**	0.8737	**0.7339**	0.7868
Decision Tree									
	Complete Case Analysis	0.9550	0.9555	0.9550	0.9547	0.8068	0.8339	0.8068	0.8193
	Zero Imputation	0.9548	0.9553	0.9548	0.9547	**0.8221**	0.8463	**0.8221**	**0.8333**
	Mean Imputation	0.9547	0.9552	0.9547	0.9545	0.8193	0.8457	0.8193	0.8314
	KNN Imputation	**0.9564**	**0.9569**	**0.9564**	**0.9562**	0.8205	**0.8468**	0.8205	0.8326
	MissForest	0.9547	0.9552	0.9547	0.9545	0.8194	0.8463	0.8194	0.8318
Random Forest									
	Complete Case Analysis	0.9798	0.9804	0.9798	0.9797	0.8444	0.8653	0.8444	0.8532
	Zero Imputation	0.9781	0.9787	0.9781	0.9781	**0.8543**	**0.8755**	**0.8543**	**0.8631**
	Mean Imputation	0.9783	0.9789	0.9783	0.9782	0.8538	0.8739	0.8538	0.8622
	KNN Imputation	**0.9804**	**0.9810**	**0.9804**	**0.9803**	**0.8543**	0.8743	**0.8543**	0.8627
	MissForest	0.9783	0.9789	0.9783	0.9782	0.8538	0.8739	0.8538	0.8622
LightGBM									
	Complete Case Analysis	**0.8683**	**0.8652**	**0.8683**	**0.8633**	0.8581	0.8620	0.8581	0.8577
	Zero Imputation	0.8531	0.8493	0.8531	0.8483	0.8657	0.8699	0.8657	0.8653
	Mean Imputation	0.8532	0.8494	0.8532	0.8483	**0.8659**	**0.8703**	**0.8659**	**0.8657**
	KNN Imputation	0.8452	0.8410	0.8452	0.8400	0.8639	0.8678	0.8639	0.8640
	MissForest	0.8532	0.8494	0.8532	0.8483	0.8624	0.8681	0.8624	0.8634
XGBoost									
	Complete Case Analysis	**0.8822**	**0.8803**	**0.8822**	**0.8787**	0.8585	0.8622	0.8585	0.8593
	Zero Imputation	0.8703	0.8677	0.8703	0.8668	**0.8659**	**0.8726**	**0.8659**	**0.8681**
	Mean Imputation	0.8713	0.8687	0.8713	0.8676	0.8647	0.8709	0.8647	0.8668
	KNN Imputation	0.8629	0.8601	0.8629	0.8588	0.8632	0.8708	0.8632	0.8661
	MissForest	0.8713	0.8687	0.8713	0.8676	0.8647	0.8709	0.8647	0.8668

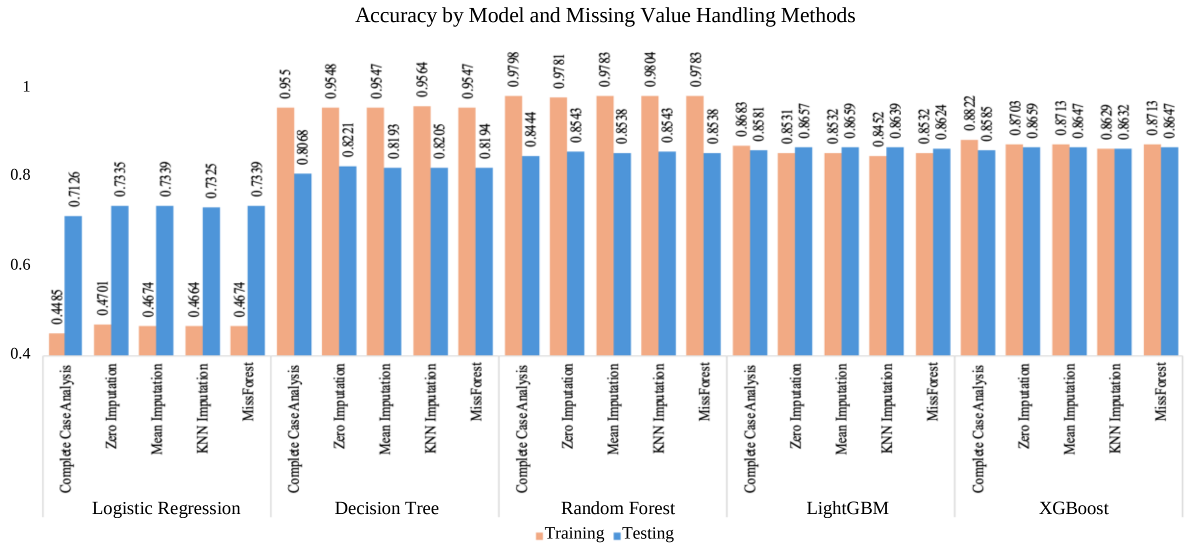

Figure 2 Accuracy across ML models and imputation methods

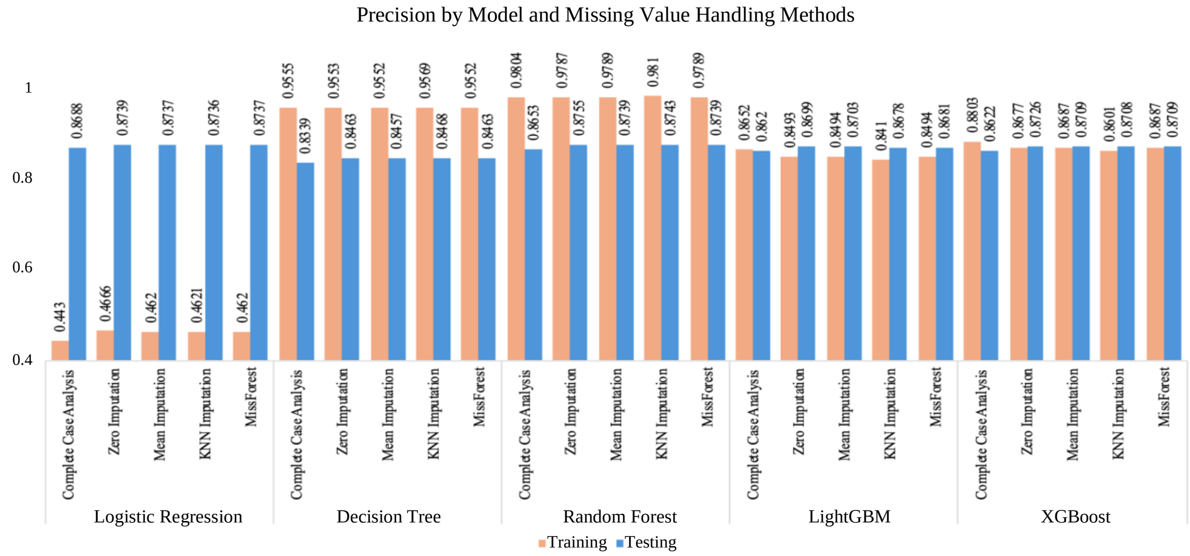

Figure 3 Precision across ML models and imputation methods

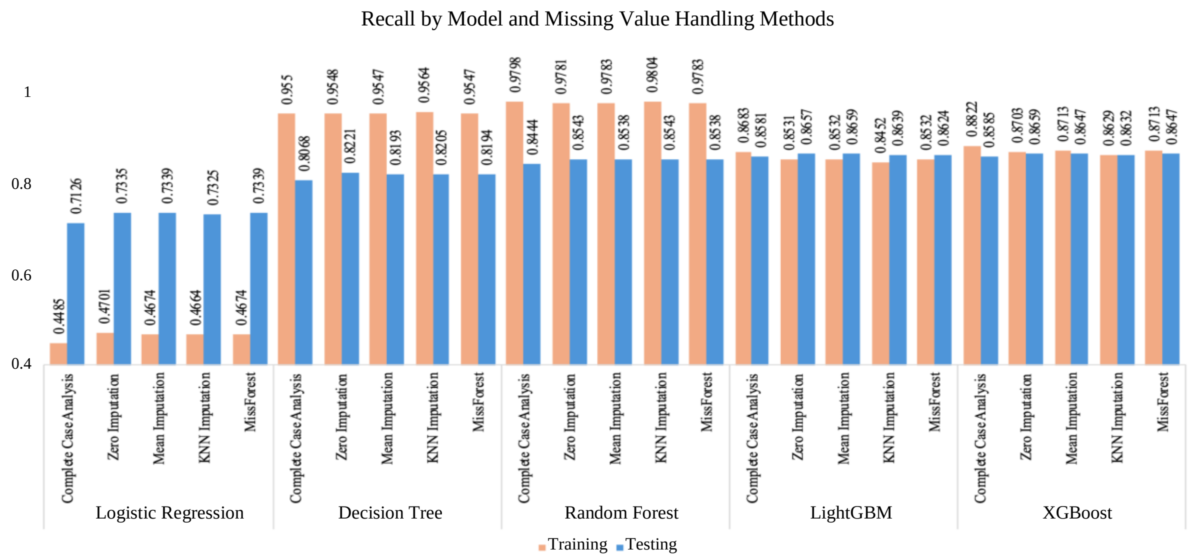

Figure 4 Recall across ML models and imputation methods

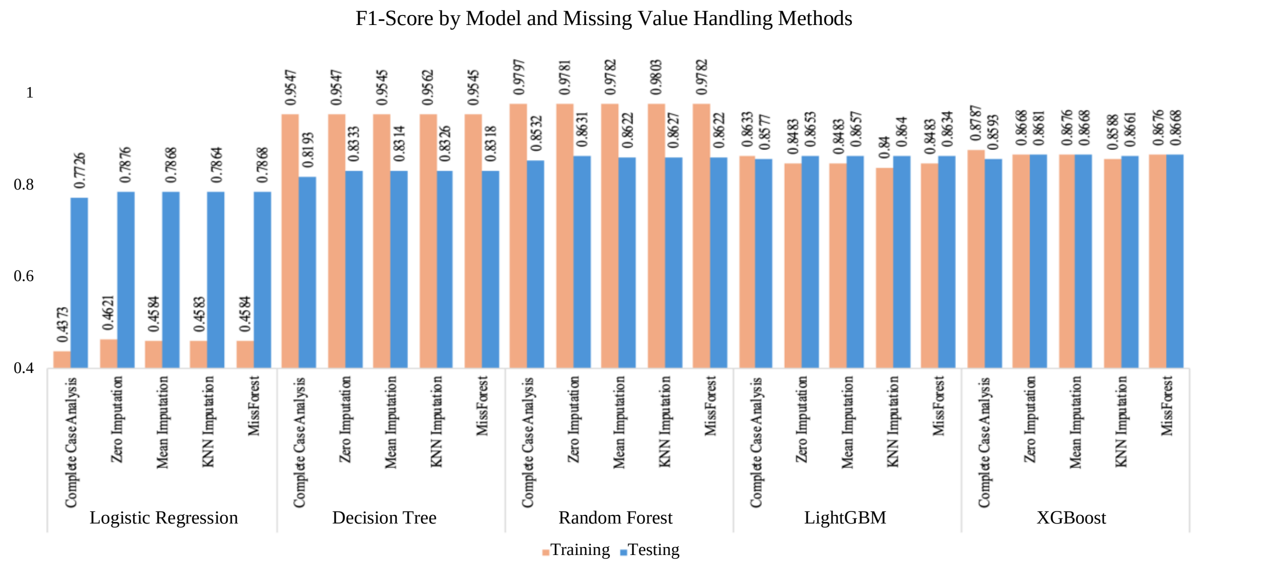

Figure 5 F1-score across ML models and imputation methods

4.2 One-vs-Rest (OvR) AUC *analysis of models and missing value handling methods*

As shown in Table 3, the AUC values of the random forest model and the k-nearest neighbor (kNN) imputation method demonstrates the eight prediction targets, including death in a day, three days, seven days, 28 days, 60 days, ICU, admit, and others, are 0.9587, 0.8245, 0.8044, 0.8818, 0.7439, 0.9337, 0.9120 and 0.9375, respectively. In contrast, when considering the highest performing model on the testing dataset, combination of XGBoost and zero imputation, it can be seen the AUC values in all of the outcomes which are 0.9947, 0.9216, 0.8848, 0.9213, 0.9533, 0.9385, 0.9119 and 0.9420, respectively.

Table 3 AUC results of ML model and missing value handling method

Model	**Method**	**AUC**							
		Death in					**ICU**	**Admit**	**Other**
		1 day	**3 days**	**7 days**	**28 days**	**60 days**			
Logistic Regression									
	Complete Case Analysis	0.9576	0.9075	0.8783	0.9096	0.9398	0.8785	0.7789	0.9160
	Zero Imputation	0.9887	0.9278	0.8956	0.9178	0.9414	**0.8876**	0.7919	**0.9206**
	Mean Imputation	0.9900	**0.9284**	**0.9031**	0.9181	**0.9475**	0.8773	**0.7934**	0.9196
	KNN Imputation	**0.9904**	0.9258	0.8908	**0.9203**	0.9459	0.8796	0.7926	0.9201
	MissForest	0.9900	**0.9284**	**0.9031**	0.9181	**0.9475**	0.8773	**0.7934**	0.9196
Decision Tree									
	Complete Case Analysis	0.5998	**0.5631**	0.5110	0.5295	0.4995	0.6258	0.7004	0.7808
	Zero Imputation	0.7917	0.5314	0.5079	**0.5298**	**0.5133**	0.6314	**0.7068**	0.7867
	Mean Imputation	0.7917	0.5069	0.5079	0.5098	0.4993	**0.6328**	0.7034	0.7858
	KNN Imputation	**0.8046**	0.5230	**0.5253**	0.5176	0.5132	0.6321	0.7044	**0.7881**
	MissForest	0.7917	0.5068	0.5167	0.5178	0.4993	0.6319	0.7039	0.7877
Random Forest									
	Complete Case Analysis	0.7439	**0.8455**	**0.8270**	**0.8837**	0.7353	0.9264	0.9089	0.9343
	Zero Imputation	0.9522	0.8187	0.8216	0.8545	0.6894	0.9317	**0.9130**	**0.9383**
	Mean Imputation	**0.9653**	0.8345	0.7811	0.8575	0.6349	0.9326	0.9113	0.9368
	KNN Imputation	0.9587	0.8245	0.8044	0.8818	**0.7439**	**0.9337**	0.9120	0.9375
	MissForest	0.9651	0.8334	0.8050	0.8596	0.7021	0.9317	0.9119	0.9366
LightGBM									
	Complete Case Analysis	0.9093	0.9087	0.9045	0.9294	0.9385	0.9332	0.9066	0.9390
	Zero Imputation	0.9565	0.9183	**0.9125**	0.9297	0.9545	**0.9404**	**0.9101**	**0.9413**
	Mean Imputation	**0.9935**	**0.9417**	0.9111	0.9254	**0.9602**	0.9376	0.9085	0.9404
	KNN Imputation	0.9551	0.9309	0.9081	**0.9332**	0.9489	0.9372	0.9073	0.9391
	MissForest	0.9816	0.9395	0.9103	0.9283	0.9536	0.9375	0.9053	0.9380
XGBoost									
	Complete Case Analysis	0.8866	0.8930	**0.8945**	0.9075	0.9227	0.9284	0.9100	0.9392
	Zero Imputation	0.9947	0.9216	0.8848	0.9213	**0.9533**	**0.9385**	**0.9119**	**0.9420**
	Mean Imputation	0.9882	**0.9274**	0.8854	0.9207	0.9432	0.9364	0.9111	0.9417
	KNN Imputation	**0.9952**	0.9154	0.8750	**0.9230**	0.9401	0.9345	0.9091	0.9402
	MissForest	0.9882	**0.9274**	0.8854	0.9207	0.9432	0.9364	0.9111	0.9417

5. Discussion

5.1 Addressing the missing data challenge in mortality prediction

A notable finding was that even basic imputation methods substantially improved linear model performance compared to complete case analysis, suggesting that in clinical settings, the utilization of any reasonable imputation strategy may be preferable to excluding incomplete records when working with linear models. As stated in Bartlett et al. [14], despite the simplicity of complete case analysis, the result of performance is under-performed other methods in this circumstance.

In contrast, tree-based models (e.g., decision tree, random forest, LightGBM, XGBoost) demonstrated greater stability across various imputation methods, likely due to their non-parametric nature and capacity to address missingness implicitly through feature splits. This stability underscores their suitability for real-world ED applications where missing data patterns are inherently unpredictable.

5.2 Advancement in multi-class mortality prediction outcomes

Due to the limitation of previous research, most of ML models for ED mortality prediction focus on a wide range of mortality outcomes which are death in seven and 30 days [10, 12] which may fail to capture the more complex patient outcomes. This research expands the predicting outcomes into eight different trajectories including death in a day, three days, seven days, 28 days, 60 days, ICU admission, general admission, and others.

The expanded range of patient outcomes allows ED to identify patients at emergent risk of death, recognize moderate-risk patient, allocate ICU resources more efficiently and reduce unnecessary hospital admissions. With all of the benefits mentioned above, it leads to higher patient satisfaction, lower operating cost and better hospital’s resources allocation.

5.3 Research implications, limitations and future directions

The research provides the findings that missing value handling can significantly improve the performance of prediction in both training and hold-out dataset. These results suggest that such a combination can effectively guide triage processes, enhancing patient care and streamlining hospital operations. The resilience of tree-based models, like random forest, across different imputation techniques highlights their potential as reliable tools in real-world ED settings, where data inconsistencies are common.

Although the research is giving promising results, it has some limitations. The data derived only from a hospital in the northeastern part of Thailand, so it might not show how patients, medical practice, or data problems are taken in other places which may restrict the application of model elsewhere. In addition, the multi-class prediction framework relies on limited sample sizes for some outcomes (e.g., only 182 cases of death within 60 days), which could lower the model’s reliability for rarer events, despite the use of SMOTE to balance classes.

In the future, researchers may consider evaluating the model in various healthcare settings, examining the role of mechanisms for handling missing data, and refining the approach to facilitate faster computation, making it suitable for real-time emergency department applications. This research utilized the original SMOTE technique to handle class imbalance. Future work can explore alternative or more advanced variants, such as Borderline-SMOTE, SMOTE-ENN, or wavelet-based ensemble approaches [29], which may offer additional performance gains in highly skewed datasets. Additionally, exploring advanced techniques, such as hybrid imputation methods or neural network-based solutions, could potentially enhance performance, particularly in datasets with substantial missing data or complex missingness patterns.

6. Conclusion

This research contributes the understanding of impact of missing value handling methods in an emergency department mortality prediction and hospitalization. The employment of XGBoost and zero imputation has been demonstrated to yield higher performance metrics in comparison to alternative methodologies with the results of accuracy, precision, recall and f1-score of 0.8659, 0.8726, 0.8659, and 0.8681, respectively, and the AUC values of death in a day, three days, seven days, 28 days, 60 days, ICU, admit, and others, are 0.9947, 0.9216, 0.8848, 0.9213, 0.9533, 0.9385, 0.9119 and 0.9420, respectively, thereby substantiating the efficacy of integrating missing value handling method with machine learning algorithms. Moreover, tree-based models have shown stability across a range of missing value approaches, a property that renders them particularly well-suited for real-world applications characterized by the presence of missing data.

The findings of this research have practical implications for healthcare institutions that are developing mortality prediction systems. The clear advantage of certain model-imputation combinations provides concrete guidance for implementation choices. Research opportunities in the future will include examining the effectiveness of these methods in different healthcare contexts and investigating performance with varying patterns of missing data.

7. Acknowledgements

This research work was supported by the Research Fund of the Faculty of Engineering, Khon Kaen University under the Research Scholarship for M.Eng. Students project under Contract Nos. M-Eng.-IE.-002/2567, Research Program, Khon Kaen University, Thailand, and BCG Economy and Sustainable Development Network through Center of Excellence Consortium under the Reinventing University System/Visiting Professor Program 2023.

8. References

[1] El Ariss AB, Kijpaisalratana N, Ahmed S, Yuan J, Coleska A, Marshall A, et al. Development and validation of a machine learning framework for improved resource allocation in the emergency department. Am J Emerg Med. 2024;84:141-8.

[2] Brossard C, Goetz C, Catoire P, Cipolat L, Guyeux C, Jardine CG, et al. Predicting emergency department admissions using a machine-learning algorithm: a proof of concept with retrospective study. BMC Emerg Med. 2025;25:1-11.

[3] Abatal A, Mzili M, Benlalia Z, Khallouki H, Mzili T, Billah MEK, et al. Hybrid long short-term memory and decision tree model for optimizing patient volume predictions in emergency departments. Int J Electr Comput Eng. 2025;15(1):669-76.

[4] VanderPlas J. Python data science handbook: essential tools for working with data. Santa Rosa: O'Reilly Media; 2016.

[5] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-30.

[6] Chang YH, Lin YC, Huang FW, Chen DM, Chung YT, Chen WK, et al. Using machine learning and natural language processing in triage for prediction of clinical disposition in the emergency department. BMC Emerg Med. 2024;24:237.

[7] Batista GEPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. Appl Artif Intell. 2003;17(5-6):519-33.
[8] Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O. A survey on missing data in machine learning. J Big Data. 2021;8:140.
[9] Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Ann Emerg Med. 2018;71(5):565-574.e2.
[10] Brasen CL, Andersen ES, Madsen JB, Hastrup J, Christensen H, Andersen DP, et al. Machine learning in diagnostic support in medical emergency departments. Sci Rep. 2024;14:17889.
[11] Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019;23(1):64.
[12] Graham B, Bond R, Quinn M, Mulvenna M. Using data mining to predict hospital admissions from the emergency department. IEEE Access. 2018;6:10458-69.
[13] Little R, Rubin D. Statistical analysis with missing data. 3rd Edition. Hoboken: John Wiley & Sons; 2019.
[14] Bartlett JW, Carpenter JR, Tilling K, Vansteelandt S. Improving upon the efficiency of complete case analysis when covariates are MNAR. Biostatistics. 2014;15(4):719-30.
[15] Goto T, Camargo CA Jr, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning–based prediction of clinical outcomes for children during emergency department triage. JAMA Netw Open. 2019;2(1):e186937.
[16] Quesada JA, Orozco-Beltran D. Analysis of missing data in electronic health records of people with diabetes in primary care in Spain: a population-based cohort study. Int J Med Inform. 2025;194:105722.
[17] Chen Z, Tan S, Chajewska U, Rudin C, Caruna R. Missing values and imputation in healthcare data: can interpretable machine learning help?. Proc Mach Learn Res. 2023;209:86-99.
[18] Psychogyios K, Ilias L, Askounis D. Comparison of missing data imputation methods using the Framingham heart study dataset. 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI); 2022 Sep 27-30; Ioannina, Greece. USA: IEEE; 2022. p. 1-5.
[19] Nagarajan G, Dhinesh Babu LD. A hybrid of whale optimization and late acceptance hill climbing based imputation to enhance classification performance in electronic health records. J Biomed Inform. 2019;94:103190.
[20] Almasinejad P, Golabpour A, Meybodi MRM, Mirzaie K, Khosravi A. A dynamic model for imputing missing medical data: a multiobjective particle swarm optimization algorithm. J Healthc Eng. 2021;2021:1203726.
[21] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002;16(1):321-57.
[22] Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. Bioinformatics. 2001;17(6):520-5.
[23] Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28(1):112-8.
[24] Nababan AA, Sutarman, Zarlis M, Nababan EB. Multiclass logistic regression classification with PCA for imbalanced medical datasets. Math Model Eng Probl. 2024;11(9):2377-87.
[25] Mienye ID, Jere N. A survey of decision trees: concepts, algorithms, and applications. IEEE Access. 2024;12:86716-27.
[26] Breiman L. Random forests. Machine learning. 2001;45:5-32.
[27] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: von Luxburg U, Guyon I, Bengio S, Wallach H, Fergus R, editors. Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4-9; Long Beach: Curran Associates Inc; 2017. p. 3149-57
[28] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Krishnapuram B, Shah M, editors. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, USA. New York: Association for Computing Machinery; 2016. p. 785-94.
[29] Majhi B, Kashyap A. Wavelet based ensemble models for early mortality prediction using imbalance icu big data. Smart Health. 2023;28:100374.