International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 11, Issue 1 (2025)
https://doi.org/10.31436/ijpcc.v11i1.561

Examining Mortality Risk Prediction Using Machine Learning in Heart Failure Patients

Mohammad Khalid Hussain*, Sharyar Wani, Adamu Abubakar

Department of Computer Science, International Islamic University Malaysia, Kuala Lumpur, Malaysia

*Corresponding author mohammad.khalid.hussain@gmail.com
(Received: 26th December, 2024; Accepted: 15th January, 2025; Published on-line: 30th January, 2025)

Abstract— Heart failure is fatal. Signs and symptoms of heart failure often overlap with those of other medical conditions. These symptoms could kill the patient. Predicting heart failure mortality helps healthcare workers spend resources to reduce or prevent deaths. Demographics, laboratory tests, and vital signs were used to create and test prediction models. This study compares random forests, and support vector machine to determine the best mortality risk prediction approach. This study analyses heart failure symptoms to identify risk factors for mortality. The study also examines how these findings apply to all heart failure patients. The study collects a subset of MIMIC-III heart failure patients to achieve this goal. Previous research studies used a smaller dataset, which is compared to this one. The experimental examination of blood creatinine, ejection fraction, and binned age shows that machine learning is be able to classify heart failure patients by mortality risk. This information helps clinicians improve treatment, improving patient outcomes and resource allocation. The study shows that machine learning can improve heart failure mortality risk prediction by using large clinical datasets like MIMIC-III. This study advances predictive analytics in healthcare, giving valuable information for clinicians and academics seeking to better heart failure patient care.

Keywords— Blood creatinine, Ejection fraction, Logistic regression, random forests, gradient boosting, heart failure

I. Introduction

Machine learning prediction tasks typically involve datasets that can be organized to distinguish between expected and unexpected outcomes, enabling effective model training and evaluation [1-2]. There exist studies suggesting that the assessment of a cardiac patient's mortality can be accomplished by evaluating the heart's ejection fraction and blood creatinine levels [3]. Nevertheless, it should be noted that the datasets employed in those studies are constrained in their scope. The dataset exhibits a significant likelihood that the model generated is not applicable to various types of patient data. The act of selecting and organizing a dataset from a publicly accessible database, such as the Medical Information Mart for Intensive Care (MIMIC-III) and subsequently implementing a comparable procedure on this dataset facilitates the development of a more comprehensive model. The dataset possesses a substantial amount of information and offers the opportunity to even incorporate race as a variable while constructing the model if need be.

Heart failure (HF) with preserved ejection fraction is common and is associated with substantial morbidity and mortality [4]. Clinical trials do not provide sufficient evidence to help guide renal dysfunction. This is despite the fact that renal dysfunction is extremely common in patients with chronic heart failure (CHF) and related to worse outcomes a lot of the treatments for CHF are the cause of worsening renal function [3]. Serum creatinine levels are an indicator of proper renal function. Serum creatinine is a waste product caused by muscle wear and tear and the kidneys are responsible for removing it from the blood. The European Society of Cardiology raises the breakoff point to greater than 50% or if its level reaches a limit of 266 µmol/L [6]. At the same time, it should be noted that the initial rise of serum creatinine may not necessarily corelate to intrinsic kidney injury but to a change in haemodynamics. This is because HF patients usually have reduced renal function at baseline and thus even a small decrease in renal function will increase serum creatinine to an extent that may even require stopping necessary medication for recovery.

The past couple of decades has seen a shift from manual record systems to electronic record systems in many fields. Information systems refer to all elements required to transfer information and respective processing procedures within an organization. Such information is collected, transformed, and disseminated by a combination of people, hardware and software, communication networks, and data resources [7].

Information systems are used in many fields such as business, economics, law, education, government administration, and medicine. Specific examples include

enterprise planning, customer relationship management, and supply chain management [7]. Information systems are extremely beneficial in areas where data is of great importance. One such area is the medical field. Application of information systems in the medical field include patient monitoring and Electronic Health Record (EHR) systems. Patient records are required for medical and legal reasons. The use of computerized patient records has been around for about four decades [8]. Computerized patient records ease documentation and administrative procedures [9]. Examples include prevention of redundant tests and patient history inquiry at point of care. The former will not only save time but also costs. In addition, they can be used for analysis to help reduce medical errors and thus improve overall healthcare [10].

An EHR is a computerized information record composed of a person's health data. It accumulates longitudinal, cross-institutional, and multi-modal health data. It also has to be based on a standard so that interoperability is possible Interoperability is important for cross-institutional usage as well as enabling a life-long patient history [11]. In summary, EHRs contain all medical information of a patient, both past and present.

The use of EHR systems assumes that users who need access are equipped with necessary equipment and software [11]. If the requirements are fulfilled, EHR systems offer several advantages: availability and ease of retrieval, access, and transfer of information. However, they are not without their disadvantages. EHR systems require substantial startup costs with regard to hardware and software. Users will also require training in order to use them. Furthermore, dedicated IT staff is required for system maintenance [11]. Examples of EHR systems include OpenMRS, HOSxP, WorldVistA, GNU Health, and OpenEMR.

Given the increasing need to derive actionable insights and solutions from existing electronic healthcare system data, significant knowledge can be extracted, particularly in areas such as heart attack and mortality risk prediction for heart failure patients. Machine learning offers powerful tools to address these challenges effectively. Consequently, the objective of this study is to examine mortality risk prediction in heart failure patients using machine learning techniques.

The structure of the paper is as follows: Section 1 provides an introduction and an overview of the research. Section 2 presents a review of relevant literature. Section 3 outlines the models and performance evaluation metrics employed in the study. Section 4 focuses on the analysis and presentation of the results, followed by Section 5, which provides a discussion of the findings. Finally, Section 6 concludes the research with key takeaways and implications.

II. RELATED WORK

The selected literature covers machine learning, healthcare systems, and chronic disease management. These studies span a wide spectrum of issues, from data preprocessing approaches in machine learning to specific clinical applications like heart failure and chronic renal disease. This review group papers by approach and examines methodology, performance indicators, and findings.

Data preprocessing and efficient machine learning methodologies are explained in Shen et al. [1] and Kumar et al. [2]. Shen et al. [1] reviewed large data preparation strategies and stressed their importance in machine learning model quality and reliability. Their investigation emphasized approaches such as feature selection, dimensionality reduction, and handling missing data, which are crucial for scalable big data applications. Accuracy, precision, and recall were used to assess preprocessing impacts across datasets. Kumar et al. [2] examined healthcare dataset machine learning predictive analytics issues. They discussed methods such as ensemble learning and neural networks, emphasizing the importance of balancing bias and variance. The authors used F1 score, MSE, and AUC-ROC to compare models, showing that hybrid models can provide reliable predictions.

Murphy et al. [3] and Redfield [4] examined heart failure with reduced and preserved ejection fractions as clinical entities. Murphy et al. [3] studied heart failure with reduced ejection fraction (HFrEF) etiology and treatment choices, stressing predictive analytics' mortality risk management potential. While their study did not expressly analyze machine learning models, the implications for applying advanced predictive tools in optimizing therapy were evident.

In his study on heart failure with preserved ejection fraction (HFpEF), Redfield [4] detailed clinical and diagnostic problems. Predictive models using biomarkers like ejection fraction and comorbidities improved clinical decision-making. Forbes and Gallagher [5] and Ponikowski et al. [6] studied chronic kidney disease (CKD) and cardiac failure. Forbes and Gallagher [5] gave CKD assessment and management a framework that stressed early detection and personalized management. Although not explicitly linked to machine learning, their findings align with predictive analytics in early disease detection.

Acute and chronic heart failure clinical recommendations were developed by Ponikowski et al. [6]. These guidelines provide structured data for machine learning models to predict patient outcomes and optimize therapy routes. Model performance could include survival rates, predicted accuracy, and rehospitalization reduction. Several researches examined healthcare infrastructure and EHRs.

https://doi.org/10.31436/ijpcc.v11i1.561

Xu and Quaddus [7] focused on information systems infrastructure management and the need for robust data storage and processing for predictive analytics.

Kirch [11] offered a succinct summary of the importance of EHRs in public health, noting their potential to enable large-scale machine learning applications. Haux et al. [10] expected healthcare information system improvements to enhance clinical decision-making and research. Giere [8], Hollerbach et al. [9], and Kirch [11] examined electronic patient information and document security. They stress the importance of secure, interoperable, and accessible data for advanced analytics and machine learning.

These researches agree that machine learning and electronic healthcare systems can transform predictive analytics, especially chronic illness management. Clinical studies give organized data and practical applications, while data preparation and infrastructure studies demonstrate the technical foundations of effective analytics. These studies measure accuracy, precision, recall, AUC-ROC, and clinical outcomes like survival and hospitalization reduction. These metrics underline the effectiveness of machine learning models in addressing healthcare challenges.

Chicco and Jurman [12] also carried out research in the same domain as Ahmad et al. [13] using the latter's dataset. However, their research differs in that they apply machine learning classifiers in contrast to purely statistical methods. They also rank the features that are significant in determining mortality risk among heart failure patients corresponding to the most important risk factors. In order to evaluate the ranking, they also perform biostatistics tests and compare them. Both methods used for feature ranking showed ejection fraction and serum creatinine as the most relevant features that affected mortality and thus, they built their mortality risk machine learning models based on these two factors alone. They learned that not only are ejection fraction and serum creatinine enough for predicting mortality risk, using just these two features for mortality risk prediction classifiers resulted in more accurate models. Their approach effectively showed that machine learning can be used for binary classification (presence of death event or not) of electronic health records of patients with heart failure.

The evaluated papers show that powerful machine learning, robust data preparation, and healthcare applications work together. They show that predictive analytics can improve clinical decision-making in heart failure and CKD care. In line with these methods, this study uses machine learning to predict heart failure mortality risk.

III. RESEARCH METHODOLOGY

The research uses the design science research methodology. Design science is used to study the "creation of artifacts and their embedding in our physical, psychological, economic, social and virtual environments." Good design improves life through the creation of "innovative, sustainable products and services," by creating value, and by mitigating any unintentional negative results of technology use. Design science combines both analysis and synthesis in product and system design by drawing from several scientific disciplines [14].

A. Data Acquisition and Preparation

The dataset, called the Medical Information Mart for Intensive Care v3 (MIMIC-III) by Johnson et al. [15] was obtained for this research [15]. The data set is a collection of 40 tables with a total of 534 columns and 728,556,685 rows organized as a relational database. In order to prepare data sets for use in the experiments, the required data had to be curated from the database and converted into a compatible format such as CSV before it could be used in the experiment. All the required features could be directly extracted from the database with the exception of the ejection fraction. This had to be specifically extracted from the clinical notes. The interim CSV files created with the previous step were passed through a filter written in the R language to extract the required ejection fraction values from the patients' echodiagram notes.

The raw dataset was generated as a result of first: it comes from patients that were diagnosed with heart failure. Patients are older than 16, i.e. patients that are not neonates. Serum creatinine and ejection fraction categories were obtained from different tables linked with the admissions table. First, the subset of patients with heart failure diagnosis was obtained by limiting the International Classification of Diseases (ICD-9) codes on the pattern 428%. This pattern helps include congestive heart failure (428.0), left heart failure (428.1), systolic heart failure (428.2), diastolic heart failure (428.3), combined systolic and diastolic heart failure (428.4), as well as general heart failure (428.9). Patients' ages and whether they died during their admission was obtained from the admission data. The former was extracted from dates of birth. Patients who were not admitted were omitted since we are focusing on inpatients.

The ejection fraction data and serum creatinine were held in different tables. Serum creatinine was obtained from the 'labevents' table. This value is regularly obtained from patients. Since we are dealing with a prediction problem, we need to take the first reading of serum creatinine when the patient is admitted.

The ejection fraction data is stored as plain text in the echodiagram data. These notes are stored in the 'noteevents' table under the category 'Echo'. Similar to serum creatinine, only the very first reading of the ejection fraction is required. But unlike serum creatinine, ejection

https://doi.org/10.31436/ijpcc.v11i1.561

fraction values need to be extracted from the echodiagram notes.

B. The Trained Models

1) Random Forest

Random forest, similar to bagging and boosting, is a binary tree-based ensemble method classifier. Ensemble methods create models with a lower variance by combining the output of multiple simpler models (often called "learners"). In order to understand the random forest ensemble method, it is imperative that decision trees are discussed. The main parts of a decision tree are nodes and branches. To make understanding easier, tree-based terminology will be discussed briefly:

- Nodes: these are of three types: (1) root/decision/parent nodes: these represents a data sample and a decision rule (e.g. patients above the age of 45?), (2) internal/chance nodes: these present the possible choices at a juncture in a decision tree that is not the root or leaf nodes of the decision tree (e.g. patients age below or equal to 45), and (3) leaf/end nodes: these are terminal nodes of a decision tree. which represent the final result of a decision tree after going through all the possible decisions at a subset of root nodes.
- Branches: these represent all possible outcomes from a decision tree, binary or otherwise. They are presented by the pathway from root nodes to leaf nodes going through internal nodes.
- Splitting: data samples at parent nodes need to be split into purer internal until they reach leaf nodes of for the respective target variable. Both continuous and discrete input variables can be used. The parent node is split into two at subsequent internal nodes into two (in the case of binary trees) "bins" depending on their value. The split is determined based on the degree of "purity" of resultant child nodes from the parent node where the split is made. This "purity" may be determined using entropy, Gini index, classification error, information gain, gain ratio, or twoing criteria. Splitting continues under the required homogeneity in the initial data sample is reached [16].

Since random forests are "binary" tree-based, the decision trees are always split into two at the nodes. At each node, the decision rule is determined by choosing a conditional rule that ensures the highest information gain. This means that the rule should provide the best separation between the available values at that level.

A random forest is a classifier consisting of a collection of tree-structured classifiers $h(x,\omega_k),k=1,...$ where the ω_k are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x.

Of course, all the trees used in a random forest cannot be giving the same classification results, since that would defeat the purpose of having the classifier in the first place. This is because if the average of the same classification results would be equal to each single classification result individually. Thus, one extra regulatory step is added at each split of every tree: only a random chosen subset of the predictors is considered. Essentially, the trees are actively prohibited from taking into consideration some of the predictors in each split. At first glance, this may seem counterintuitive, but it makes sense once we realize that we are aiming for a classification model that be used on a generalized data set, thus this will provide the model a greater perception of the training data. This is what results in a more robust classification model.

2) Support vector machine (SVM)

Support vector machine (SVM) is a supervised learning classifier used in classification problems. SVM can be explained simply if we consider a plot of two groups of multiple points on a 2D coordinate plane. SVM will try to determine the best line to create a split between the two groups to help classify any new points. A basic classification model would attempt to place a straight line between the points where the differentiation is clear, i.e. it is easily determined which groups a certain coordinate point is from. The same would apply in a 3D plane but the only difference is that in this case, SVMs will try to determine the best plane instead of line to classify the training points. At the heart of SVMs is an optimization problem. This problem takes a function, e.g. the equation of a line in a clear 2D plane situation and determines the best values that ensure the maximum distance between points of the two groups that are closest to the split; these points are known as support vectors. The functions that are optimized can be changed depending on the situation of points on a plane.

C. Experimental Analysis Tools

The MIMIC-III database was originally created using PostgreSQL and was thus uploaded to a local schema instance of PostgreSQL; there is no choice of choosing a different RDBMS. Jupyter notebook was used to extract the required data. Python was used as the programming language. It was used in a Google Colaboratory (Colab) environment. SQL queries were written within strings in a Python context and then executed to generate results. The ejection fraction of patients in the MIMIC-III database were not always available as values in a column. Instead, some records of ejection fraction were stored in the patients' clinical notes. In order to extract them from the clinical records, an R language filter created by Major [17] was used.

https://doi.org/10.31436/ijpcc.v11i1.561

The filter extracts the ejection fraction values from the clinical notes using regular expressions (Regex).

In order to provide comma-separated values (CSV) files for the Jupyter notebooks, the patient data was first extracted from the PostgreSQL database and saved as CSV files. These CSV files were then processed using the aforementioned R filter by Major [17]. Once the ejection fraction values were extracted, the required sub-datasets for the different experiments were created using Jupyter notebook and processed. The sets of features used to predict mortality are {serum creatinine (SC), ejection fraction (EF)}, {serum creatinine (SC), ejection fraction (EF), age}, and {serum creatinine (SC), ejection fraction (EF), age, sex}.

IV. PRESENTATION OF THE RESULTS

The research paper examining mortality risk prediction using machine learning in heart failure patients presents an investigation of a number of different machine learning models that were applied to the MIMIC-III dataset. Some of the most important variables, including serum creatinine (SC), ejection fraction (EF), and age, are included in the dataset. These variables are analyzed according to several demographic categories, including overall, men, and females. The results of the evaluations of models with SC and EF are presented in Table I. A poor area under the curve (AUC) of 0.4931 indicates that the Random Forest Classifier (RFC) has limited predictive capacity. However, it attained a moderate level of accuracy (72.14%). The fact that the model had low recall and F1 scores demonstrates that it is unable to accurately identify situations that are actual affirmative responses. With an accuracy of 76.50% but an area under the curve (AUC) of 0.0, the SVM with Linear Kernel performed incredibly badly, indicating that there was no distinction between the classes. There was a lack of satisfaction with both the recall (10.89%) and the F1 scores (4.14%).

TABLE I
. RESULTS OF MIMIC-III DATASET WITH SC AND EF

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Random Forest Classifier	0.7214	0.4931	0.1631	0.1504	0.1558	-0.0100	-0.0101	0.821
SVM - Linear Kernel	0.7650	0.0000	0.1089	0.0558	0.0414	-0.0049	-0.0107	0.061

Results using SC, EF, and Age are included in Table II, which demonstrates some small improvements. RFC demonstrated an increase in accuracy (77.45%) and area under the curve (0.5404), indicating a somewhat enhanced capacity to differentiate between outcomes. There was a noticeable improvement in the predicted reliability, as evidenced by the fact that the precision reached 20.53%. A higher recall of 43.27 percent was demonstrated using SVM, but it lacked precision and F1 score consistency, which indicated that the predictions were not balanced. The incorporation of age results in a marginal improvement in prediction power, particularly for RFC. Nevertheless, the models, on the whole, have difficulty delivering reliable forecasts, which indicates that there is a requirement for either additional features or other techniques.

TABLE II
RESULTS OF MIMIC-III DATASET WITH SC, EF, AND AGE

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Random Forest Classifier	0.7745	0.5404	0.1597	0.2053	0.1790	0.0510	0.0517	0.865
SVM - Linear Kernel	0.5453	0.0000	0.4327	0.0984	0.1336	-0.0052	0.0011	0.068

As compared to SVM, RFC produced a higher accuracy (73.32%) and area under the curve (AUC) (0.5741) when SC and EF were used (see Table III) but only for male patients. However, its inadequate capacity to detect good outcomes is highlighted by its recall rate of 19.84% and its F1 score of 18.28%. However, SVM had a worse F1 score (18.86%) and a lower precision (11.42%), despite having a better recall (64.47%).

TABLE III
RESULTS OF MIMIC-III DATASET WITH SC AND EF FOR MALES ONLY

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
SVM - Linear Kernel	0.4165	0.0000	0.6447	0.1142	0.1886	0.0111	0.0159	0.1610
Random Forest Classifier	0.7332	0.5741	0.1984	0.1730	0.1828	0.0261	0.0264	0.7460

RFC attained a high level of accuracy (81.29%) with SC and EF (Table IV) when limited to female patients, however its area under the curve (AUC) was 52.97%. Poor true positive predictions are indicated by the recall rate (10.03%) and the F1 rate (14.47%). With a recall of 0.37% and a precision of 3.33%, SVM performed worse.

https://doi.org/10.31436/ijpcc.v11i1.561

TABLE IV
RESULTS OF MIMIC-III DATASET WITH SC AND EF FOR FEMALES ONLY

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
SVM - Linear Kernel	0.8388	0.0000	0.0037	0.0333	0.0067	-0.0016	-0.0012	0.038
Random Forest Classifier	0.8129	0.5297	0.1003	0.2675	0.1447	0.0626	0.0733	0.642

There was a decrease in performance for both models when Age (Table V) was included for the males-only dataset (previously, only SC and EF were used). The RFC scored higher in accuracy (82.75%), but it had a significantly lower recall rate (3.57%). Each and every metric revealed that SVM performed poorly.

TABLE V
RESULTS OF MIMIC-III DATASET WITH SC, EF, AND AGE FOR MALES ONLY

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Random Forest Classifier	0.8275	0.5445	0.0357	0.1925	0.0589	0.0042	0.0118	0.551
SVM - Linear Kernel	0.8184	0.0000	0.0393	0.0141	0.0208	-0.0064	-0.0111	0.036

Although there were some slight gains for RFC when Age was taken into account (Table VI) for the females-only dataset, metrics like as recall (6.37%) and F1 (9.91%) remained at a low level. When it came to all measures, SVM struggled. Differences between the sexes are revealed by the findings. In general, RFC performs better than other methods; yet its recall and F1 scores continue to be low, particularly for females. This indicates that there is a requirement for individualized models or additional features in order to handle variances that are specific to demographics.

TABLE VI
RESULTS OF MIMIC-III DATASET WITH SC, EF, AND AGE FOR FEMALES ONLY

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Random Forest Classifier	0.8206	0.5248	0.0637	0.2333	0.0991	0.0346	0.0440	0.563
SVM - Linear Kernel	0.7712	0.0000	0.1000	0.0153	0.0265	-0.0057	-0.0145	0.021

V. Discussion

In terms of area under the curve (AUC), the Random Forest Classifier consistently performed better than SVM, which indicates that it is better able to discriminate. However, memory and F1 scores reveal a challenge in recognizing true positives, particularly for females and when additional variables (such as age) are included in the analysis. SVM with Linear Kernel showed low performance across both AUC and F1 scores in every scenario. This was the case regardless of the situation. In spite of the fact that recall was higher in certain instances, precision and overall dependability were not satisfactory. In this particular setting, the RFC model is more trustworthy than the SVM model for predicting mortality risk; nonetheless, both models need to be optimized. Bad recall and F1 scores point to an excessive reliance on majority class predictions, which indicates that there are issues associated with an imbalanced dataset.

As a result of the consequences of the study, it is clear that the inclusion of Age in the feature importance makes a marginal improvement in predictive skills, but it is not sufficient on its own. Additional clinical characteristics, such as comorbidities and biomarkers, have the potential to improve the performance of the model. Considering that the performance of male and female patients differed significantly from one another, the findings underscore the necessity of developing gender-specific predicting techniques. When compared to SVM, RFC displays superior overall performance; however, due to its limits in detecting positive cases, it is necessary to further refine or investigate ensemble approaches and deep learning. The findings of this study highlight the potential of machine learning to assist in the process of clinical decision-making for patients suffering from heart failure. Existing models, on the other hand, need to be improved in order to guarantee their dependability and generalizability in real-world situations. Inadequate recall and F1 scores are indicative of datasets that are not balanced or hyperparameters that are not ideal. It is of the utmost importance to address these challenges by implementation of strategies such as oversampling or advanced model tweaking. The results of this study shed insight on the potential benefits and difficulties associated with using machine learning to forecast mortality risk in individuals suffering from heart failure. Despite the fact that RFC demonstrates potential, the low recall and F1 scores highlight the necessity of enhancing feature engineering, gender-specific modeling, and advanced methodologies in order to increase prediction accuracy and clinical relevance

VI. Conclusions

The study, "Examining Mortality Risk Prediction Using Machine Learning in Heart Failure Patients," addresses the critical challenge of improving mortality risk prediction in heart failure patients using Serum Creatinine (SC), Ejection

Fraction (EF), and Age as predictive variables. Heart failure is a leading cause of mortality globally, yet traditional risk assessment methods often lack precision in identifying high-risk individuals. The study aimed to evaluate the performance of machine learning models, specifically the Random Forest Classifier (RFC) and Support Vector Machine (SVM), in predicting mortality risk based on the MIMIC-III dataset. The findings revealed that while RFC outperformed SVM across most metrics, including accuracy and area under the curve (AUC), both models struggled with recall and F1 scores, indicating difficulty in accurately predicting true positive cases. Gender-specific analyses highlighted disparities in model performance, with female-only data exhibiting lower recall and predictive reliability, underscoring the need for tailored approaches. The inclusion of Age improved predictive performance slightly, but results were still suboptimal, pointing to the need for richer datasets and advanced feature engineering. The study has significant implications for clinical practice, emphasizing the potential of machine learning to augment traditional risk stratification in heart failure care. However, the findings also highlight limitations in current models, such as imbalanced data handling and insufficient feature representation, which restrict their clinical applicability. Future research should focus on integrating additional clinical variables, exploring deep learning methods, and addressing data imbalances to enhance model accuracy and generalizability. Moreover, gender-specific modeling should be prioritized to reduce disparities in predictive outcomes. Overall, this research contributes to the growing field of AI-driven healthcare by demonstrating the promise and challenges of machine learning in mortality risk prediction for heart failure patients, paving the way for more effective, personalized interventions.

ACKNOWLEDGMENT

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

REFERENCES

[1] H. Shen, W. Ma, and Y. Wang, "A review on data preprocessing techniques for machine learning in big data era," Frontiers of Computer Science, vol. 17, no. 2, pp. 163–182, 2023. doi: 10.1007/s11704-023-10123-6.

[2] A. Kumar, N. Goyal, and D. Singh, "Efficient prediction using machine learning techniques: A systematic review of challenges and methodologies," Applied Intelligence, vol. 52, no. 7, pp. 7284–7304, 2022. doi: 10.1007/s10489-021-02742-1.

[3] S. P. Murphy, N. E. Ibrahim, and J. L. Januzzi, "Heart failure with reduced ejection fraction: A review," JAMA, vol. 324, no. 5, pp. 488–504, 2020.

[4] M. M. Redfield, "Heart failure with preserved ejection fraction," New England Journal of Medicine, vol. 375, no. 19, pp. 1868–1877, 2016.

[5] A. Forbes and H. Gallagher, "Chronic kidney disease in adults: Assessment and management," Clinical Medicine, vol. 20, no. 2, p. 128, 2020.

[6] Ponikowski, P., Voors, A. A., Anker, S. D., Bueno, H., Cleland, J. G., Coats, A. J., … others. (2016). 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Kardiologia Polska (Polish Heart Journal), 74(10), 1037–1147.

[7] J. Xu and M. Quaddus, "Managing Infrastructure for Information Systems," in Managing Information Systems: Ten Essential Topics, Paris: Atlantis Press, 2013, pp. 85–107. doi: 10.2991/978-94-91216-89-3_6.

[8] W. Giere, "Electronic patient information–pioneers and MuchMore," Methods of Information in Medicine, vol. 43, no. 5, pp. 543–552, 2004.

[9] A. Hollerbach, R. Brandner, A. Bess, P. Schmücker, and B. Bergh, "Electronically signed documents in health care," Methods of Information in Medicine, vol. 44, no. 4, pp. 520–527, 2005.

[10] R. Haux, E. Ammenwerth, W. Herzog, and P. Knaup, "Health care in the information society. A prognosis for the year 2013," International Journal of Medical Informatics, vol. 66, no. 1–3, pp. 3–21, 2002.

[11] W. Kirch, Ed., "Electronic Health Record (EHR)," in Encyclopedia of Public Health, Dordrecht: Springer Netherlands, 2008, pp. 326–326. doi: 10.1007/978-1-4020-5614-7_946.

[12] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Medical Informatics and Decision Making, vol. 20, no. 1, pp. 1–16, 2020.

[13] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: A case study," PLoS One, vol. 12, no. 7, p. e0181001, 2017.

[14] P. Y. Papalambros, "Design science: Why, what, and how," Design Science, vol. 1, p. e1, 2015. doi: 10.1017/dsj.2015.1.

[15] A. E. Johnson, T. J. Pollard, L. Shen, L.-W. H. Li, M. Feng, M. Ghassemi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Scientific Data, vol. 3, p. 160035, 2016.

[16] N. Patel and S. Upadhyay, "Study of various decision tree pruning methods with their empirical comparison in WEKA," International Journal of Computer Applications, vol. 60, no. 12, 2012.

[17] AMIA Annual Symposium Proceedings, vol. 2016, p. 844, 2016, American Medical Informatics Association