International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Evaluation of Machine Learning Methods for Predicting Heart Failure Readmissions: A Comparative Analysis

[1]Samit Shivadekar, [2]Ketan Shahapure, [3]Shivam Vibhute, [4]Ashley Dunn

Submitted: 24/09/2023 **Revised**: 17/11/2023 **Accepted**: 28/11/2023

Abstract: A heart failure (HF) condition is a type of chronic cardiovascular disease that affects millions of people globally. It can lead to various symptoms and has a significant impact on the quality of life. Despite the advancements that have been made in treating this condition, it remains a major public health issue. One of the biggest challenges that HF management faces is the high number of readmissions. This issue contributes to the increasing of patients' outcomes and costs the healthcare system. Implementing effective interventions and identifying those at high risk of returning to the hospital can help lower the financial burden on the system. Through the use of machine learning techniques, researchers can now predict the likelihood of HF readmissions. These tools can analyze large datasets and provide a personalized diagnosis and treatment plan. There have been various studies that have examined the use of ML for predicting HF readmissions. The goal of this study is to analyze the various techniques used in predicting HF readmissions and provide a comprehensive analysis of their performance. Through a combination of data collected from various sources, including a diverse set of patients, we will be able to explore the performance of various ML algorithms. In addition to the algorithms' performance, we will also look into their impact on various parameters, such as model evaluation metrics, optimization techniques, and feature selection. The findings of this study will be used to inform policymakers and healthcare providers about the use of ML techniques to identify patients at high risk of HF readmissions. These insights can help them improve the quality of care for those with this condition and develop effective interventions. The objective of this study is to use the power of ML to improve the management of HF and reduce the burden of readmissions on both the patients and the healthcare systems.

Keywords*: *Regression, SVM, Heart Disease, Classifier, ML algorithms

1. Introduction

One of the most common causes of death among patients is a heart attack. The way to treat cardiovascular illness is by analyzing vast amounts of data. This data can be used to anticipate, monitor, and treat various infections, such as coronary episodes. Large data examination is a process commonly used in the corporate world to control, distinguish, and oversee massive amounts of information. It can be used to anticipate, analyze, and treat cardiovascular diseases. Through machine learning and data mining, various tools were used to analyze and extract information from the data [13].

The prevalence of heart disease is one of the factors that contribute to the development of hazardous complications. It has been estimated that around 26 million people worldwide have this condition. Every day, there is a rise in cardiovascular problems[1]. One of the most important factors that can be considered when it comes to addressing the increasing number of cardiovascular problems is the development of a machine learning program that can classify and categorize the data. This will allow the clinic to improve its efficiency and prevent heart failure[1].

The goal of our work is to improve the accuracy of the data collected by the Digitalis Investigations Group (DIG) for detecting heart disease. Through a variety of computer learning methods, we were able to analyze the data and estimate the likelihood of heart failure in a clinical library. The findings of our study suggest that the use of machine learning techniques in predicting heart failure and other conditions could be useful in analyzing live patient data. We utilized a framework that included five different algorithms. The analysis demonstrates major progress and high precision relative to prior studies. Relative to the latest research, the data collection must be improved with respect to the analyzed data sample. The limited size of the data collection is the key constraint of this research. The data collection included a small amount of health documents which are expanded using different techniques. For fact, the findings revealed that the device would be effective and beneficial in prompt detection of heart failure in the patient for physicians and cardiac surgeons. We contrasted the utility of Random Forestry, raise, Random Fostering and Logistic Regression (LR), as

[1]*Department of CSEE, University of Maryland Baltimore County, USA*
Email id: samit1@umbc.edu
[2]*Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, USA*
Email ID: ketans1@umbc.edu
[3]*San Jose State University, San Jose, California, United States*
Email Id: shivamvinayak.vibhute@sjsu.edu
[4]*Manager, Religroup, University of Maryland, Baltimore County, USA*
Email id: ashley.dunn@religroupinc.com

well as poison regression against standard LR to forecast all-day readmissions and readmissions attributable to heart failure for 30 and 180 days through the use of surveillance details. Randomly, 50% of patients were chosen for a derivative collection and the remaining patients were tested for 100 bootstraps. This research is not meant to substitute the conventional method used to diagnose and forecast heart failure but to facilitate this phase by leveraging new technology such as machine learning (ML). The ML is not a modern methodology and has been used for different purposes for many years. [2]

Some researchers are developing AI approaches that can help identify chronic illnesses in their early stages. Wearable devices that are light and dependable are ideal for healthcare applications. They can also be used to monitor the health of people. The continuous changes in the body that happen through a few clinical mindfulness exercises can help improve a person's quality of life. Early detection and treatment of diseases are some of the most crucial factors that people can consider when it comes to improving their well-being. Large amounts of data are collected in medicinal services, which can be useful in delivering effective and efficient care.

Through wearable devices, people can get the most out of their health data and improve their quality of life. They can also help medical professionals keep track of their patients. In addition, it can help them develop effective treatment plans and recommendations for the most common illnesses. The rapid emergence and evolution of new data frameworks and techniques has created various challenges for organizations that rely on current satellite communication systems and electronic PCs. These associations should consider adopting the latest innovations to improve their position in the field[14].

2. Literature Review

K. Polaraju et al, proposed Prediction of Heart Disease utilizing Multiple Regression Model and it demonstrates that Multiple Linear Regression is proper for anticipating chances of heart disease. The work is performed utilizing preparing an informational collection consisting of 3000 instances with 13 distinct properties which was referenced before. The dataset is divided into two sections that is 70% of the information utilized for preparing or training the data and 30% utilized for testing. In view of the outcomes, plainly the classification precision of the Regression algorithm is better contrasted with other algorithms.[7]

Across the advanced and affluent world, cardiovascular disease is the main cause of death. Changes in the life pattern, social, and cultural transformation of the population residing in developed countries are significant factors for rising the CVD (cardiovascular disease) incidence. This discovery has contributed to rigorous preventive studies. Diagnosis of risk factors and CVD predictors will allow physicians to better identify patients at high risk and to avoid the disease [3].

Despite fast developments in medical technologies and testing methods today, the prior list of CVD risk factors includes further predictors. We must therefore design updated methods for risk assessment to screen people at high risk early in their lives. This research describes, classifies, and explains cardiovascular risk factors, how they function, and what approaches to avoid CVD progressions can be applied [3].

Worldwide there are more deaths from cardiovascular disorder than any other disorder [4]. Muscles and nerves of most CVDs become dysfunctional along the heart and blood supply pathways. This study analyzes the function of hypertension and cholesterol in coronary disease development and progression at various densities of triglycerides. Within this analysis we will also analyze potential biomarkers, such as homocysteine, fibrinogens, D-dimer and composites of thrombin / antithrombin III, interleukin and serum amyloid [4].

Diagnosing cardiovascular disorder is a challenging job and may provide an automatic diagnosis of patient cardiac failure to promote more therapy. As a result, the diagnosis of cardiovascular diseases has received huge global interest in healthcare. In diagnosing heart disease with enhanced effectiveness, artificial intelligence has played an important role. Many latest reviews in literature were carried out from this point of view. The analysis of these diagnostic techniques can therefore lead to further developments in this sector [4].

Abdur Razzak et al., submitted a detailed survey of 47 articles published from 2005 to 2013 in the standard journal providing a summary of cardiovascular disorders, as well as a study on several types of cardiovascular conditions such as coronary heart attack, pulmonary artery disease, heart failure, respiratory diseases, renal disorder, heart valve disease, and hypo plastic left pulmonary syndrome. The study showed that approaches focused on the neural network lead to greater efficacy and certain strategies have reached a greater precision of 90 percent. Ultimately, the analysis question is approached in the same way to further the previous studies findings [5].

Seema et al. utilized various methods such as the Decision tree, SVM, and AI to analyze the data gathered from healthcare records. They were able to predict sickness with high accuracy. A relative analysis is performed to determine the best execution of classifiers on a given target. For instance, SVM performed better than Nave Bayes when it came to accuracy when it came to predicting diabetes[9].

Chala Beyene et al. the use of data mining techniques could help predict the occurrence of heart disease. The

objective is to analyze and interpret the data in a short time to find out the condition's early symptoms. The proposed method is similar to the one utilized in medicinal services by analyzing the data collected by specialists who have no additional information. It can detect if a patient has a coronary illness[12].

3. Selected Data Set

The DIG Trial is a multi-center, randomized, double-blind, placebo-controlled study that is focused on assessing the safety and effectiveness of Digoxin in treating individuals with a type of heart condition known as congestive cardiovascular failure. Over 200 years ago, Digitalis was first presented to the public. It has since been widely recommended as a treatment for cardiovascular breakdown. Although few studies have shown that Digoxin can reduce the side effects of this condition, it can also improve the quality of life for patients. Unfortunately, many of the studies that were conducted on the use of Digoxin for cardiovascular breakdown were small. They focused on the effects of the drug on the patients' symptoms. They also failed to address the effects of the treatment on their cardiovascular results. Despite the small number of studies that were conducted on the use of Digoxin for cardiovascular breakdown, the harmful effects of the drug were still unknown[15].

The data collected from the primary DIG trial is used to analyze the effects of digoxin on various aspects of cardiovascular health. In the primary analysis, patients with a qualifying heart condition who had a 45% or less discharge component were randomly given either a fake treatment or placebo. The results of the analysis revealed that those who received the fake treatment had higher cardiovascular mortality rates and hospitalization due to other health conditions[15].

This program is composed of various books that are loaded with the necessary tools to perform the analysis. The first section of the program focuses on the data collection and provides a variety of visualizations. Later in the program, the algorithms are used to perform the analysis[7].

- Age (age in years)
- Sex (1 = male; 0 = female)
- CP (chest pain type)
- TRESTBPS (resting blood pressure (in mm Hg on admission to the hospital))
- CHOL (serum cholesterol in mg/dl)
- FPS (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- RESTECH (resting electrocardiographic results)
- THALACH (maximum heart rate achieved)
- EXANG (exercise induced angina (1 = yes; 0 = no))
- OLDPEAK (ST depression induced by exercise relative to rest)
- SLOPE (the slope of the peak exercise ST segment)
- CA (number of major vessels (0-3) colored by fluoroscopy)
- THAL (3 = normal; 6 = fixed defect; 7 = reversible defect)
- TARGET (1 or 0)

Fig 1.1 Selected dataset

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

Fig 1.2 Data model

Fig 1.3 Null values

4. Methodology

To classify people with heart disease and healthier citizens, the proposed program is created. Specific computer teaching models have been evaluated on full and selected functions to assess cardiac attack diagnosis. Component discovery algorithms like Relief, mRMR (Minimum redundancy maximum relevance feature selection) and LASSO (regression analysis method) have been used to pick critical features and classifier output has been evaluated on these chosen apps. The Evidence Collection for Cardiac Failure in Cleveland has been applied and is included in many studies [11]. Logistic regression K-NN, ANN, SVM, DT and NB have been used in the framework for the common machine learning classificatory. The reliability and success measurement parameters of the model have been determined. The methodology of the proposed framework was organized in five separate steps including (1) data set preprocessing, (2) collection of features, (3) cross validation procedure, (4) object classification, and (5) performance evaluation methodologies of the classifier. The structure for the current method is shown in Figure 1.

We evaluated the efficacy of wild-forests, boosting, random forests hierarchically paired with vector supports or logistic regression (LR), and Poisson's regression against conventional LR, utilizing data from the Monitoring to Increase Heart Failure Outcomes study to forecast all-time heart disease readmissions and reactions during 30-and 180-day cycles. We randomly selected 50%

of patients, and the rest of the patients were validated with 100 bootstrapping for a derivative set. We contrasted C data on inequality in risk deciles for predictive spectrum and distributions of observed outcomes [7, 8].

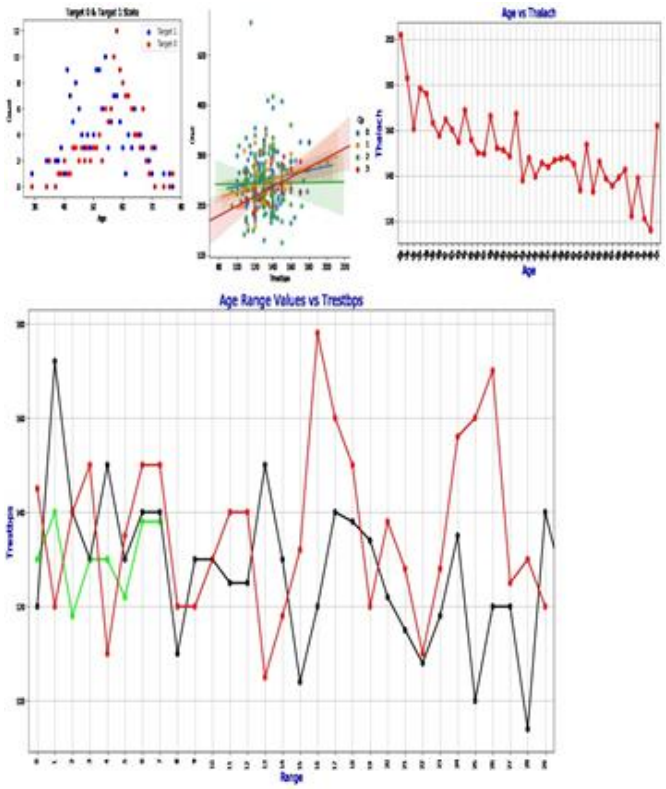
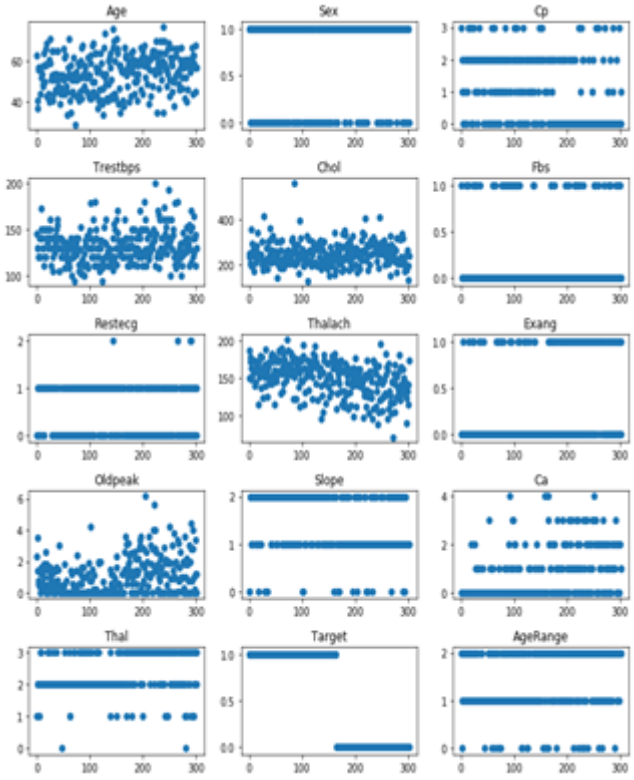


Fig 1.4 Result statistics

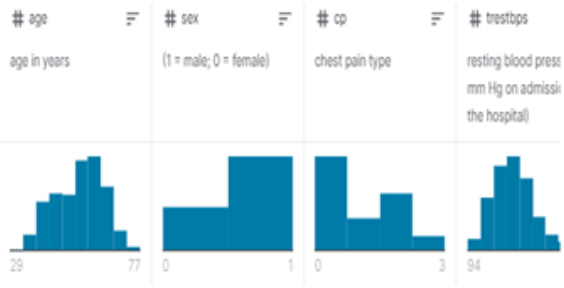
The highest performing computer education pattern, Random Forests, increased by 17.8% relative to LR during the 30-day all-round readmission prediction (mean C figures, respectively 0.628 and 0.533). Of readmissions owing to cardiac insufficiency, C results increased by 24.9% relative to LR (mean C estimates of 0.678 and 0.543, respectively). The observed re-entry rates for 30-day all-cause readmission showed a significantly wider range than LR (14.2 percent, 16.4 percent), in both the lowest and highest deciles of predicted risk with random forest (7.8 and 26.2 percent, respectively) [8].



As is established, much of our data is not just classified by those characteristic values. To reduce these conditions, we must carry out size operations. Most activities of this scope are ongoing. That is as it looks [10,11].

- ➢ Normalization
- ➢ Z-score

Within a data collection, a certain scaling is performed to measure data that is distinct from each other. The data is 0.1 because of this process. Any scaling operation can alter that. In our service the normal and standardization scales are used. Between the data obtained there is a great change. We must therefore use this SVM algorithm method. The previous parts mention all correlation values between the results. Consequently, these assets are utilized in multiple areas across numerous activities. The purpose of this listing is to ensure that the p-value method establishes a hypothesis and an inference according to a conclusion submitted in each function. Throughout this method, the relationships of all other properties are analyzed after deciding the class property as a hypothesis. This gives every property a specific amount. The crucial point here is that such figures are not even 1.00. If the number is near to 1.00 then we can conclude that our results are not good [10].



5. Conclusions

The prediction of readmission for heart insufficiency following hospitalization was enhanced by machine learning approaches in contrast to LR and given the highest predictor range in observation duration.

The analysis of the treatment of basic cardiac details helps save human lives and identify irregularities of cardiac problems at an early point. Machine learning methods have been used for the analysis of raw data and for offering fresh and groundbreaking insight into heart disease. Prediction of heart failure in the medical sector is complicated yet very necessary. Nevertheless, if the illness is identified at early stages and aggressive intervention is taken as quickly as possible, the death risk may be significantly reduced. Another expansion of this thesis is strongly recommended, rather than only abstract methods and models, to apply analysis to actual data sets. The suggested HRFLM (Hybrid Random Forest with Linear Model) method incorporates Random Forest (RF) and Linear Model (Logistic regression) characteristics. For diagnosis of heart attack, HRFLM proved very successful. Diverse combinations of machine learning techniques for improved predicting techniques will pursue the future course of this study. To achieve a broader perception of the significant characteristics to improve heart diseases prediction, new feature selection methods may also be created [1, 7, 10].

Classifiers performance evaluation metrics

Predictive model	Turning parameters	Accuracy (%)	Specificity (%)	Sensitivity (%)	MCC	AUC (%)	Processing time (s)
Logistic regression	$C=1$	88	98	76	88	87	16.213
	$C=10$	87	98	76	88	87	16.200
	$C=100$	89	98	77	89	88	16.111
	$C=0.001$	74	98	47	72	73	16.233
K-nearest neighbor	$K=1$	80	73	78	80	80	24.400
	$K=3$	75	80	72	76	76	24.500
	$K=7$	74	78	71	75	75	24.600
	$K=9$	73	78	70	75	73	24.611
	$K=13$	70	69	71	70	71	21.777
Artificial neural network	16	77	2	100	50	69	21.600
	20	54	96	5	50	68	22.101
SVM (kernel = RBF)	$C=100, g=0.0001$	87	95	78	86	87	14.134
	$C=1, g=0.01$	79	82	81	79	80	14.139
	$C=10, g=0.001$	75	84	68	76	77	14.255
SVM (kernel = linear)	$C=10, g=0.0001$	78	95	55	78	74	18.139
	$C=100, g=0.0001$	80	97	60	79	79	18.222

References

[1] Micheline Kamber and Jian Pei Jiawei Han, Data Mining Concepts and Techniques, 2012.

[2] Mohammed Abdul Khaled, Sateesh Kumar Pradhan and G. N. Dash, "A survey of data mining techniques on medical data for finding locally frequent diseases", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, pp. 149-153, August 2013.

[3] Christian Nordqvist. “Heart Disease: Definition, Causes, Research.” Medical News Today, April 7, 2016. http://www.medicalnewstoday.com/articles/237191.php.

[4] Porimol Palma,"Heart diseases now a plague in Bangladesh",[online]Available: http://www.thedailystar.net/heart-diseases-now-a-plague-in-bangladesh-57098.December 2014

[5] Abdur Razzak, "A questionnaire survey on infectious disease among hospital patients in Kushtia and Jhenaidah Bangladesh", International Journal of Genetics and Molecular Biology, vol. 3, no. 9, pp. 120-xxx, October 2011.

[6] D. P. Shukla, Shamsher Bahadur Patel, and Ashish Kumar Sen, "A literature review in health informatics using data mining techniques", International Journal of Software & Hardware Research in Engineering, vol. 2, no. 2, February 2014.

[7] V. Gayathri, M. Chanda Mona and S. Banu Chitra, "A survey of data mining techniques on medical diagnosis and research", International Journal of Data Engineering (IJDE) Singapore Journal of Scientific Research (SJSR), vol. 6, pp. 301-310, 2014.

[8] M. Akhil Jabbar, Priti Chandra, and B. L Deekshatulu, "Heartdisease prediction system using associative classification and genetic algorithm", International Conference on Emerging Trends in Electrical Electronics and Communication Technologies-ICECIT, 2012.

[9] R. Chitra and V. Seenivasagam, "Review of heart disease prediction system using data mining and hybrid intelligent", ICTACT Journal on Soft Computing, vol. 03, no. 04, July 2013.

[10] Abhishek Taneja, "Heart disease prediction system using data mining techniques", Oriental Journal of Computer Science & Technology, vol. 6, pp. 457-466, December 2013.

[11] Hlaudi Daniel Masethe and Mosima Anna Masethe, “Prediction of heart disease using classification algorithms", Proceeding of the World Congress on Engineering and Computer Science, vol. II, 2014.

[12] H.D.UCI,” kral machine/analyzing-the-heart-disease /notebook, Heart Disease UCI,

2019."[Online].Available: https://www.kaggle.com/kralmachine/analyzing-the-heart-disease/notebook. [Accessed 2020].

[13] Ahmed Ismail, Samir Abdlerazek, I.M. El-Henawy, "Big data analysis in heart disease prediction",JATIT,15th June 2020. Vol.98. No 11

[14] Adil Hussain Seh, Dr Pawan kumar Chaurasia,"A Review on Heart Disease Prediction Using Machine Learning Techniques ",IJMRA,Vol. 9 Issue 4, April 2019, ISSN: 2249-0558,

[15] NHLBI," Digital Investigation Group (DIG) data Documentation" NHLBI Study #995 Revised Feb 1992 https://biolincc.nhlbi.nih.gov/media/teachingstudies/digdoc.pdf?link_time=2020-08-21_12:31:27.479822

[16] Dhabliya, D., Sharma, R. Cloud computing based mobile devices for distributed computing (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 1-4.

[17] Dhabliya, D., Parvez, A. Protocol and its benefits for secure shell (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 19-23.