Vishav Mehra, 2025, 13:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Analysis of Cardiovascular Diseases with Causes and Future

Vishav Mehra
Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat

Abstract- A complicated clinical disease known as heart failure manifests as a variety of symptoms and indicators. Even with improvements in our knowledge of its pathogenesis and the availability of several treatment options, it continues to be a major source of morbidity and death. Precise risk assessment is necessary to direct managerial choices and support patient-focused treatment. While recognized risk variables have historically been the focus of epidemiological studies to predict mortality in heart failure, newer machine learning algorithms provide a unique method for uncovering other important predictors. In this review, we examine the role of supervised learning algorithms in predicting mortality in heart failure, with a view to facilitating the development of evidence-based treatment guidelines and health care policies.

Keywords- Cardiac problem, morbidity, Epidemiological studies, Supervised learning algorithms

I. INTRODUCTION

Around 31% of the fatalities caused all around the world are caused due to cardiovascular diseases (CVD's) making them the leading cause of mortality. Heart failure is frequently caused due to cardiovascular diseases and 12 components make up this dataset, which may be used to calculate mortality through cardiovascular breakdown. Most of the CVD's can be mitigated by applying the methodologies that benefit the whole population by watching out for the amicable danger factors such as usage of tobacco, inadequate nutrition, gaining excessive weight and drinking of excessive alcoholic beverages that cause problems. Individuals who have a heart infection or are suffering from a cardiac ailment (usually because of the presence of ailments like high blood pressure, diabetes or ailments that were diagnosed before) require early affirmation and treatment where an AI model can be of astounding assistance [1].

Various assessments have tended to plan the fatality and diligence of patients with Heart failure looking at changed time- frames in various nations. According to the Framingham Heart Study, the decline rate after stopping Heart failure in the United States was roughly 10% after 30 days (about 4 and a half weeks), 20%-30% after a year, and 45-60% after 5 years. Of course, the Rotterdam trial, which looked at HF patients across Europe, found a reduced death rate, with 11 percent and 41% passing rates after one year and five years, respectively [2].

Serum creatinine and ejection fraction are two crucial markers that are frequently measured in the diagnosis and management of heart failure. Different cut-off values for ejection fraction have been used to define heart failure in epidemiological research. For example, the Glasgow study used a cut-off of 30%, while most other studies have used higher cut-offs of 40-45% [3]. The continuous existence of heart failure seems to be surging among studies despite these categorization discrepancies, suggesting that the cut-off value selection may not have a substantial impact on the overall burden of the illness. More research is necessary to determine the underlying causes of this consistency, which are yet unknown [4].

Despite advances in medical understanding and treatment choices, epidemiological studies have revealed a disturbing trend in recent years of an increase in heart failure hospitalizations and a rise in fatality rates. This demonstrates the critical need for efficient methods to assess the severity of the disease and forecast death in heart failure patients. The development and implementation of novel medications, interventions, and healthcare policies require accurate prediction models that account for various factors such as age, gender, co- morbidities, and lifestyle factors. Early detection of coronary disease is critical in modern society, where it is becoming increasingly prevalent due to lifestyle changes and demographic shifts. A reliable cardiovascular disease prediction model would improve patient outcomes and save healthcare costs by reducing the need for costly treatments and hospital stays[5]. This becomes a suggested approach for early detection of coronary heart disease.:

- In order to create predictions and compare outcomes, the dataset contains 12 distinct parameters, including age, anaemia, creatinine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum creatinine, serum, and mortality outcome (0 or 1).
- This is a binary classification dataset so various supervised learning algorithms have been utilized such as decision tree, random forest, Ada boost classifier to anticipate the fatal event.

Our study is organized in such a way that in section 1 we introduced the readers with the alarming issue of mortality due to heart failure. In section 2 we presented the related work and in section 3 we presented the result methodology. Probing to section 4 we presented the evaluation of methods and finally concluded our work in Section 5.

II. LITERATURE REVIEW

Available to the public on Kaggle and UCI data warehouses (Kaggle.com), the material focuses on decision making. Using the clinical records of the patients, prediction models were created to identify those who were at risk of cardiovascular disease and to forecast the severity of the condition. Data from 194 men and 105 women are included in the collection; 299 of the patients have been diagnosed with cardiovascular disease. Numerous clinical factors, including age, gender, blood pressure, and smoking history, have been thought to be predictive of a patient's risk for cardiovascular disease (CVD). Data is pre-processed and cleaned to ensure the quality of data before deploying in predictive models. To develop such models surges the assistance in direct clinical judgements and enhances the outcomes of the patients in:.

- **Age -** It shows the age in years. The range is 40 and 95 years while the typical age is 60 years.
- **Pallor -** It shows the events the individual has encountered sickliness. Expecting that the haematocrit levels of an individual are lower than 36% then the center specialist considered a patient is having anemia. The range is in the center and 1.00 while the typical stands at 0.43.
- **Creatinine phosphokinase -** It searches for the CPK compound level in blood. CPK compound streams into the blood due to hurt in muscle tissues. By virtue of which the extended level CPK in the blood of patients can provoke cardiovascular breakdown or some injury [68]. The arrive at tests from 23.000000 and completing at 7861.000000 and the typical is 581.83964.
- **Diabetes -** It goes from 0.00 to 1.00 and the typical stands at 0.418060.
- **Ejection fraction -** The release division quality surveys and processes how much blood the left ventricle guides out with each pressure in rate. It is at 14.000000 and 80.000000 and the ordinary is 38.083612.
- **High Blood Pressure -** This trademark tends to the diastolic heartbeat in mm Hg. It has an ordinary of 0.351171.
- **Platelets-** It tells the incorporate of the platelets present in the body. It is from 25100.000000 to 850000.000000 while the ordinary lies at 263358.029264.

- **Serum creatinine-**The serum creatinine is delivered by creatine, it is a kind of waste which gets delivered when a muscle isolates. Serum creatinine is especially drawn in by experts to truly investigate the limit of the kidney.
- **Serum sodium-**The serum sodium test is a standard routine-based blood test which shows the level of sodium present in blood. Cardiovascular breakdown might be the clarification for the weird decline in the level of sodium present in blood.
- **Sex -** It exhibits the direction of the patient encountering cardiovascular breakdown.
- **Death event (result)-** The death event incorporates or the target in our equal plan study, states whether the encountering patient died or made due before the completion of the ensuing time frame, that was 130 days on ordinary .

Estimations used in assumption for Heart failure prediction:

- **Extra Trees**-It is a social event of AI estimations that relates the gauges from a couple of decision trees. It can consistently instate as-incredible or favored execution over the estimation of sporadic forest, yet it uses an accommodating computation to build the decision trees which is used as a piece of the outfit [6]. The maximum accuracy achieved for Extra Trees is 0.8342.
- **AdaBoost classifier-** It is a meta-assessor that starts its connection by fitting an authentic classifier on the certified dataset and a while later fits the additional copies of the classifier on the identical dataset in which the proper classifier is implanted at this point where the presence of incorrectly requested events are worked with so much that the inconvenient cases get drawn in by the resulting classifiers [7]. Highest accuracy achieved for ada boost classifier is 0.7866.
- **Naïve Bayes:** This grouping of estimations relies upon Bayes speculation. It has free affirmation for the components. It's a prohibitive probability model, which contemplates the specific responsibility from every part, ignoring the relationship between the components. It needs a little dataset to arrange the classes for getting ready and this is a fundamental advantage of this particular computation. It achieved the highest accuracy of 0.8314 [8].
- Linear Discriminant Analysis is a dimensionality need strategy that is routinely used for request issues which are coordinated. It is used for showing openings in packs that are disengaging something like two classes[9]. It is used to expand the features into a lower perspective space from higher viewpoint space. 0.8542 is recorded to be Linear Discriminant Analysis’s highest accuracy.
- Cox regression (or proportional hazards regression) is one of the regression models used in medical research for drawing the relationship between the survival time of victims and the predictor variables. Investigation of severity of several factors in a particular time upon a specified event that is going to take place is conducted by Cox regression[10]. Binary events (in which the output is either 0 or 1) such as death use Cox regression for survival analysis.
- XG Boost algorithm- A well-known gradient boosting approach is termed as Extreme gradient boosting often called as XG Boost that escalates the execution, performance and pace of tree based machine learning algorithms.
- Logistic regression is one of the classification algorithms that are a part of Supervised learning methods and are best suited for categorial variables . Simply said, it is a machine learning model that is used to determine or predict the likelihood of a binary (yes/no) event occurring. It achieved the highest accuracy amongst all the algorithms of 0.8598.
- Decision Classifier, dataset classification is handled by internal nodes, decision rules are represented by branches, and the conclusions are offered by leaf nodes. The curacy achieved for decision classifier is 0.7819 for this dataset.

- LASSO is a regression analysis technique that incorporates variable choice with regularization to increase the predictability and understandability of a statistical model

III. PROPOSED FRAMEWORK

To carry out the classification and prediction as indicated by the method, an algorithm has been created, and Figure 1 presents a visual depiction of the results.

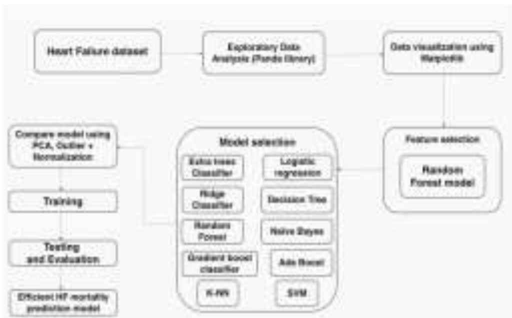

Figure1: Algorithm for classification and prediction

This study utilizes a dataset containing various factors relevant to predicting heart failure-related mortality. The dataset was loaded into a pandas data frame within the Pycaret environment using the read command. Different visualizations such as matplotlib and seaborn were used to summarize the distribution of data [11].

By identifying and using the appropriate characteristics and variables in the setup environment, we were able to evaluate models using a variety of measures, including area under the curve (AUC), accuracy, and precision.

To be employed in machine learning models, the input dataset was divided into training and testing sets. The final data was then converted. One target variable was used in an Extra Trees Regressor model. In addition, to obtain better accuracy than the prior model, a Logistic Regression model was used. Additionally, a Random Forest model and a Ridge Classifier were used. Lastly, measures were used to provide a rating and strength summary of the models' performance. T The evaluation findings of these models can help forecast heart failure-related mortality, enabling early intervention and perhaps improving patient outcomes.

IV. RESULTS

Checking the distribution of data

The distribution of the data is critical for anticipating or categorizing a problem. This will enhance the model's search for a heart disease-related trend in the data. In order to check attribute values and determine the skewness of data (asymmetry of a distribution), several distribution plots are produced [12].

The data is shown in a variety of charts so that a wide overview of the data may be studied. The distribution of age and death event, the distribution of age and ejection fraction, the distribution of serum sodium rate, and the number of deaths are analyzed, as well as the features that are important for heart disease and those that are not important for heart disease, and conclusions are drawn as shown in Figures 2, 3, 4, and 5.

Age and Death event
Scatter plot

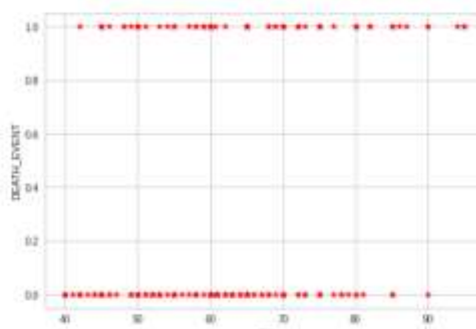

Figure 2: Scatter plot between Age and Death event

Graphs are a prominent tool that help in visualization of variables, which in return becomes more useful for humans, who can easily draw conclusions from them. A scatter plot is a group of dots plotted along axes. In statistics, scatter plots are really beneficial since they show how much, if any, relationship there is between observed quantities or occurrences (called variables).

Here, scatter plot and histogram are used to depict the number of death events in different age groups. The graphs depict that the highest number of deaths due to heart failure have been attained between 60 – 70 age group people.

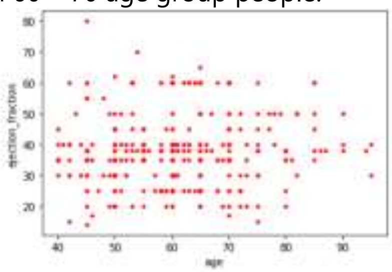

Figure 3: Scatter plot between age and ejection fraction

Age and ejection Fraction

Ejection fraction being the most prominent attribute in predicting the mortality rate due to heart failure in the previous results show a scatter plot for no correlation (It is a type of scatter plot with unclear increasing or decreasing trend) [13].

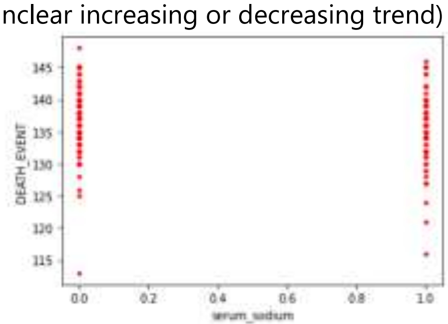

Figure 4: Scatter plot between Death events and quantity of serum sodium

Death event and serum sodium

Death event being a binary target variable, that is having an output of only 0 or 1(whether the person dies due to heart failure or not). This distribution plot clearly illustrates that the ratio of quantity of serum sodium is almost equal in both the cases which shows that it does not play an important role in predicting heart failure mortality [14].

Feature Significance/importance

Through the use of just relevant entries and the elimination of noise, it is a technique for lowering the number of input variables. It entails choosing the right dependent on the kind of challenge, your model of machine learning should have you're trying to solve automatically[15].

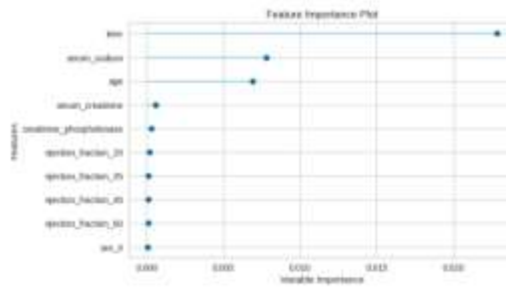

Figure 5: Feature selection plot of Logistic regression

The above plot is Feature importance plot of Logistic Regression model which shows us the result that features like (time, serum sodium ratio, age) also have a leading role in accuracy of model while the leftover parameters have a minor but significant role in achieving the accuracy.

Evaluation process used

Confusion Matrix/Error Plot (For Classification)

Confusion matrix, accuracy score, precision, recall and various compare model functions are employed in the evaluation procedure.

A type of table that contains both true and expected values is termed as Confusion matrix. It is divided into four sections: True positive is the first section, that designates the values as true and are also true in actuality. False Positive is the second section, that occurs when erroneous terms are recognized as true. False negative is designated as the third section, occurring when a value is true but is incorrectly labeled as negative [16]. The fourth option, true negative (TN), ensures that the number was indeed negative and that it was appropriately identified as such.

Figure 6 shows the confusion matrix of this dataset

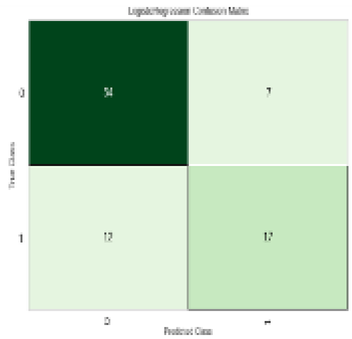

Figure 6: Confusion matrix of Logistic regression

All the predicted results of random forest model have been well concluded in confusion matrix

- **Precision-** The accuracy of a machine learning model's positive prediction is one measure of its performance and that measure is termed as the precision of the model. For this dataset the precision is .84786086
- **Accuracy-** Out of all the data points, the number of points that were successfully predicted are known as the accuracy of the model. The accuracy of this dataset is 0.81707317
- **Recall -** The measure to identify the number of True Positives accurately is termed as the recall of the model. Therefore, recall tells us how many individuals, out of all those with heart disease, we correctly identified as having it. For this dataset, the recall is 0.82978723

ROC curve

The ROC curve is a graphical representation of a binary classifier's performance that plots the TPR versus the FPR as the decision threshold changes. It allows researchers to evaluate and compare binary classifiers based on their trade-off between the TPR and FPR. The TPR is the proportion of actual positives that the model correctly identifies, while the FPR is the proportion of actual negatives that the model incorrectly identifies as positives. The area under the ROC curve (AUC) quantifies the model's performance, with higher values indicating better predictive power and lower values indicating poor performance. The ROC curve is widely used in machine learning to optimize model performance and improve decision-making[17].

The ROC curve indicates that the logistic regression model distinguishes between positive and negative outcomes in the heart failure dataset rather effectively. By analysing the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR), as indicated by the curve, researchers may evaluate the model's ability to predict true positives while lowering false positives. An ideal model would have a ROC curve that reaches the top-left corner of the graph, indicating a TPR of 1 (100%) and an FPR of 0 (0%).

The AUC score for the model is 0.8347, indicating that the overall performance of the model is reasonably good. The curve reaches near the top-right corner of the graph, suggesting that the model is capable of predicting true positives at a reasonable rate, but it is also generating some false positives. The ROC curve is a useful tool that allows researchers to evaluate the trade-off between true positives and false positives and choose an appropriate threshold for the model's predictions.

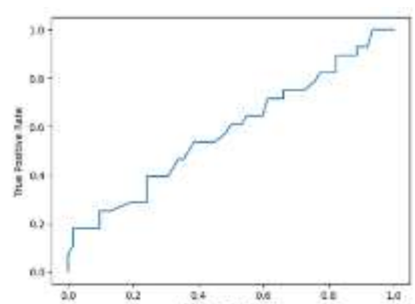

Figure 7: ROC curve based on logistic regression

Compare Model function

The compare_models() function is a useful tool in the pycaret library that enables the comparison of various machine learning models. It automates the process of creating and evaluating several models using a specified metric, such as accuracy or AUC, and ranks them based on their performance[18]. This function works with any dataset, automatically dividing it into training and testing sets, preprocessing the data, and applying various machine learning techniques to it. Researchers may rapidly determine which algorithm performs best at predicting heart failure based on certain predictor factors by using the compare_models() function with the heart failure dataset. On comparison of numerous models and selecting the most optimal one, researchers can surge the accuracy of predictions, that would lead to better results in prediction of heart failure for patients[19].

Regularly comparing the results of many machine learning algorithms is essential. The main goal of model comparison and selection is to improve machine learning applications' usefulness and performance. The aim is to determine the optimal algorithms that fulfil the demands of data and business needs [20].

The performance of ten machine learning algorithms using compare model function with Principal Component Analysis (PCA) and outlier removal + normalization have been evaluated. Table 1 compares the accuracy, precision and recall in different machine learning algorithms using compare model function and compare model function using PCA.

It can be observed that on using the compare model function, Extra trees classifier achieved the highest accuracy of 83.42% with 0.8365 of precision and 0.8000 of recall while it was followed by Ridge Classifier and Random Forest Classifier by an accuracy of 82%. The lowest accuracy of all was achieved by SVM algorithm of 62.19%. When using compare model PCA (principal component analysis), Extra trees achieved an accuracy, precision and recall of 64.5% , 0.5731 and 0.4929 respectively and was totally outshined by Ridge classifier and Linear Discriminant Analysis by achieving an accuracy of 69.82%, precision of 0.5832 and 0.6175 and recall of 0.4946. The Decision tree classifier carried out the accuracy of 53.80 % with a precision of 0.4183 which is lowest of all in both of the fields . Additionally, the compare_models() function could be useful for researchers who are inexperienced with machine learning or lack the information necessary to select the optimal approach for their dataset. The feature reduces the chance of errors during human evaluation and saves time by automating the model selection process. Numerous machine learning methods are also supported by the function, such as logistic regression, random forest classifiers, decision tree classifiers, and support vector machines [21].

By supporting multiple algorithms, the function ensures that researchers can choose the best model for their dataset regardless of the algorithm type. compare_models() also provides an option for fine-tuning the hyperparameters of the top-performing models using a grid search or a randomized search [22]. This feature enables researchers to optimize the models' performance and improve their predictions further.

Table 1: Evaluation using compare model and compare model function using PCA

Models	With compare Model function			With compare model using PCA		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Extra Trees	0.8342	0.8635	0.8000	0.6450	0.5731	0.4929
Logistic Regression	0.8147	0.8260	0.7778	0.6877	0.5832	0.4946
Ridge Classifier	0.8295	0.8499	0.7889	0.6982	0.6175	0.4946
Random Forest	0.8243	0.8380	0.7889	0.6658	0.6632	0.4893
Naïve Bayes	0.7550	0.7094	0.8556	0.6716	0.6733	0.4179
Ada Boost	0.7497	0.7570	0.7111	0.5640	0.4888	0.4625
LDA	0.8243	0.8402	0.7889	0.6982	0.6175	0.4946
K-NN	0.6266	0.6082	0.6222	0.6503	0.5961	0.4911
SVM	0.6219	0.6383	0.6111	0.5971	0.4895	0.4929
Decision Tree	0.7819	0.7805	0.7778	0.5380	0.4183	0.4643

Table 2 shows the performance analysis using compare model function (Outlier removal + normalization) when taking all feature subsets into consideration. It is evident that the Logistic regression achieved the highest accuracy and recall of 85.98% and 0.8056 respectively of any other classification model but was outshined by SVM machine learning model in precision, which stood at 0.9139 followed by LR at 0.8839 while on the other hand Decision tree achieved the lowest in all the three parameters as compared to all other machine learning models. Logistic regression was followed ridge classifier in accuracy and precision by 84.87% and 88.39.

Table 2: Compare model function (Outlier removal + Normalization)

Models	With compare Model function using Outlier + Normalization		
	Accuracy	Precision	Recall
Extra Trees	0.8376	0.8844	0.7542
Logistic Regression	0.8598	0.8839	0.8056
Ridge Classifier	0.8487	0.8876	0.7806
Random Forest	0.8258	0.8399	0.7792
Naïve Bayes	0.8314	0.8249	0.8056
Ada Boost	0.7866	0.8027	0.7176
LDA	0.8542	0.9019	0.7806
K-NN	0.8317	0.8771	0.7542
SVM	0.8484	0.9139	0.7556
Decision Tree	0.7144	0.7217	0.6944

Among all of the classifiers applied on the dataset, the Ridge classifier and Logistic Regression showed significant results for heart failure prediction with an average accuracy of 78.70 % and 79.20 % respectively but the AUC (Area under curve) of Ridge classifier, extracted using above classifiers showed poor performance making it unfit for the prediction but on the other hand the Logistic Regression has produces excellent classification performance as its AUC lies at 0.8222 that clearly shows that it is able to perform better classification [12].

V. CONCLUSION

Currently, heart failure’s disease research area has a lot of significance and it can make a patient's life much better if it is detected at an early stage. The recent research and developments have produced promising results. In our work, the problem of deaths due to Heart Failure is coped through a machine learning approach and comparing many models for machine learning enables researchers to identify the algorithm that performs the best overall. This paper shows 4 methods for evaluation and comparative analysis was performed, and encouraging results were obtained. The conclusion that is founded here is that Logistic regression suits best for the dataset and the feature selection plot based on Logistic regression shows the importance of other various parameters like sex , number of major vessels etc.

Along with Ejection fraction, traditional biostatistics studies have identified serum creatinine as the other most important factor. Furthermore, the research demonstrated that machine learning might be used to categorize people's electronic health records. with cardiovascular heart disease into binary groups. Now the limitation for the present study is that the small sized datasets are used (299 patients). More information for the physical features of patient, perhaps their employment history would've been helpful to examine other cardiovascular disease risk factors

This check yielded some acceptable findings that differ from the original dataset. Age, high blood pressure, serum creatinine, and ejection fraction are the top characteristics identified by the researchers; high blood pressure is ranked eighth out of eleven, and anemia is ranked tenth out of eleven.

This research has a good impact on medical practice, which can become a supporting tool for all the doctors for predicting whether the heat failure will cause death or not by just focusing on the ratio of serum creatinine in the body and ejection fraction.

REFERENCES

1. D. J. G. Chicco, "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, p. 16, 3 12 2020.

2. I. B. G. Bytyçi, "Mortality in heart failure patients," Anatolian Journal of Cardiology, vol. 15, no. 1, p. 63, 1 1 2015.
3. D. J. G. Chicco, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Medical Informatics and Decision Making, vol. 1, no. 1, p. 1, 3 2 2020.
4. Ford, E. S., & Capewell, S. (2007). Coronary heart disease mortality among young adults in the US from 1980 through 2002: concealed leveling of mortality rates. Journal of the American College of Cardiology, 50(22), 2128-2132.
5. Brenner, M. H. (1971). Economic changes and heart disease mortality. American Journal of Public Health, 61(3), 606-611.
6. Goldman, L., & Cook, E. F. (1984). The decline in ischemic heart disease mortality rates: an analysis of the comparative effects of medical interventions and changes in lifestyle. Annals of internal medicine, 101(6), 825-836.
7. Sidney, S., Go, A. S., Jaffe, M. G., Solomon, M. D., Ambrosy, A. P., & Rana, J. S. (2019). Association between aging of the US population and heart disease mortality from 2011 to 2017. JAMA cardiology, 4(12), 1280-1286.
8. Capewell, S., Beaglehole, R., Seddon, M., & McMurray, J. (2000). Explanation for the decline in coronary heart disease mortality rates in Auckland, New Zealand, between 1982 and 1993. Circulation, 102(13), 1511-1516.
9. Capewell, S., Beaglehole, R., Seddon, M., & McMurray, J. (2000). Explanation for the decline in coronary heart disease mortality rates in Auckland, New Zealand, between 1982 and 1993. Circulation, 102(13), 1511-1516.
10. Tunstall-Pedoe, H., Kuulasmaa, K., Mähönen, M., Tolonen, H., & Ruokokoski, E. (1999). Contribution of trends in survival and coronar y-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA Project populations. The Lancet, 353(9164), 1547-1557.
11. Bertuccio, P., Levi, F., Lucchini, F., Chatenoud, L., Bosetti, C., Negri, E., & Vecchia, C. L. (2011). Coronary heart disease and cerebrovascular disease mortality in young adults: recent trends in Europe. European Journal of Cardiovascular Prevention & Rehabilitation, 18(4), 627-634.
12. Khaw, K. T., & Barrett-Connor, E. (1987). Dietary fiber and reduced iscremic heart disease mortality rates it men and women: A 12-year prospective study. American Journal of Epidemiology, 126(6), 1093-1102.
13. Palmieri, L., Bennett, K., Giampaoli, S., & Capewell, S. (2010). Explaining the decrease in coronary heart disease mortality in Italy between 1980 and 2000. American journal of public health, 100(4), 684-692.
14. Berry, Colin, David R. Murdoch, and John JV McMurray. "Economics of chronic heart failure." European journal of heart failure 3.3 (2001): 283-291.
15. Cohn, Jay N. "The management of chronic heart failure." New England Journal of Medicine 335.7 (1996): 490-498.
16. Groenewegen, Amy, et al. "Epidemiology of heart failure." European journal of heart failure 22.8 (2020): 1342-1356.
17. Kannel, William B., and Albert J. Belanger. "Epidemiology of heart failure." American heart journal 121.3 (1991): 951-957.
18. Mudd, James O., and David A. Kass. "Tackling heart failure in the twenty-first century." Nature 451.7181 (2008): 919-928.
19. Butler, Javed, et al. "Incident heart failure prediction in the elderly: the health ABC heart failure score." Circulation: Heart Failure 1.2 (2008): 125-133.
20. Javeed, Ashir, et al. "Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: a systematic review and future directions." Computational and Mathematical Methods in Medicine 2022 (2022).
21. Maragatham, G., and Shobana Devi. "LSTM model for prediction of heart failure in big data." Journal of medical systems 43 (2019): 1-13.

22. Olsen, Cameron R., et al. "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure." American Heart Journal 229 (2020): 1-17.