



# Predicting Survival of Heart Failure Patients Using Classification Algorithms

*Oladosu Oyebisi Oladimeji, Olayanju Oladimeji*

*University of Ibadan, Ibadan, Nigeria*

## ARTICLE INFORMATION

Received: July 22<sup>nd</sup>, 2020  
 Revised: September 28<sup>th</sup>, 2020  
 Available online: September 30<sup>th</sup>, 2020

## KEYWORDS

Heart Failure, Classification Algorithms,  
 Machine learning, Data Mining

## CORRESPONDENCE

E-mail: [oladimejioladosu@gmail.com](mailto:oladimejioladosu@gmail.com)

## A B S T R A C T

Heart failure is a situation that occurs when the heart is unable to pump enough blood to meet the needs of other organs in the body. It is responsible for the annual death of approximately 17 million people worldwide. Series of studies have been done to predict heart failure survival with promising results. Hence, the purpose of this study is to increase the accuracy of previous works on predicting heart failure survival by selecting significant predictive features in order of their ranking and dealing with class imbalance in the classification dataset. In this study, we propose an integrated method using machine learning. The proposed method shows promising results as it performs better than previous works and this study confirms that dealing with imbalanced dataset properly increases accuracy of a model. The model was evaluated based on metrics such as F-measure, Precision-Recall curve and Receiver Operating Characteristic Area Under Curve. This discovery has the potential to impact on clinical practice, when health workers aim at predicting if a patient will survive heart failure. Attention may be focused on mainly serum creatinine, ejection fraction, smoking status and age.

## INTRODUCTION

It is a known fact that heart is an essential organ of human body [1]. The heart generates blood to other part of the organs in the body. Where it fails to do so or something goes wrong while performing its responsibility, it could lead to immediate death of the person or heart failure! In essence heart failure is a situation whereby the heart is unable to pump enough blood to other organs in the body. It is usually caused by conditions such as diabetes, high blood pressure, or other heart conditions or diseases like HIV, thyroid disorders, alcohol abuse, congenital disease, etc [2,3]. It is the situation that occurs when muscle in the heart wall fades and enlarges, limiting heart pumping of blood [4]. It is important to note that heart disease is one of the most popular diseases in middle aged citizens [5]. As affirmed by the World Health Organization (WHO) [6], Cardiovascular disease (CVD) which is the heart and blood vessels disorders, is now top reason for death causing 31% of deaths annually! Also, Chicco and his colleague [3] stated that it causes the death of approximately 17 million people worldwide annually!

Since the past decades, machine learning and data mining are becoming very popular in the prominent researches in virtually every aspect of human activities and the healthcare sector is not exempted. Machine learning is a subset of artificial intelligence that scientifically studies algorithms and statistical models which computer systems use in order to perform a specific task effectively without using explicit instructions, but relying on patterns and inference instead [7,8]. The healthcare industry has

accumulated big data [9]. Data analytics has been of great help for bringing insight into big data for diagnostic, disease prevention, prediction and policy-making purposes in healthcare industry [10,11].

Data analytics is advancing in healthcare industry but modelling survival for heart failure is still a problem nowadays, both in terms of achieving high prediction accuracy and identifying the driving factors [3]. Virtually all of the models developed for this purpose reach only modest accuracy [12]. More recent times models are showing promising improvements, especially if coupled with additional targets [3].

In this study, analysis was carried out on a dataset of medical records of patients having heart failure. Ahmad and his colleagues [4] in 2017, worked on this dataset using traditional biostatistics time-dependent models in order to predict mortality and identify the main features of patients having heart failure, from their medical records. Later, Zahid and his colleagues [13] in 2019, analyzed this same dataset to create two different gender-based survival models; in essence, one for men and the other for women. Although, these two works [4,13] show promising results, the problem was tackled by standard biostatistics methods; thereby leaving room for machine learning methods. This made Chicco and Jurman [3] to use machine learning methods to predict survival of patients with heart failure in which their results showed that serum creatinine and ejection fraction alone are enough to predict mortality of heart failure.

However, major challenges in model learning is the feature selection problem, as the feature selection step is so important in machine learning with the purpose of eliminating unnecessary

and unimportant features [9,14–16] and also dealing with imbalanced dataset. It is not always possible to generate a good predictive model for imbalance dataset; approaches have been proposed to this issue in which we try to use in this paper in order to address the problem.

Similarly, the main objective of the research paper is to apply machine learning algorithms to predict survival of heart failure patients by addressing the issue of imbalanced dataset which [3] did not address and feature selection in order to achieve better accuracy compared to [3]. The second section discusses the methodology used in this research, while the third section showcases the result, followed by the discussion of the result and finally the conclusion is drawn at the fourth section.

## METHOD

The description of the proposed methodology is given below:

1. Pre-processing (data manipulation and normalization)
2. Feature selection: performed using four different algorithms couple with ranker search method.
3. Classification – classifiers were tested: kNN, SVM, Naïve bayes, random forests.

4. Evaluation of results - based on confusion matrix (Accuracy, ROC AUC, PR AUC and F-measure metrics).

The proposed methodology for this work was formulated using WEKA software that is an open source software for machine learning which was developed at the University of Waikato. The dataset that was used to pinpoint this research was gotten from UCI Machine Learning Repository [17] which is heart failure clinical records dataset by [4] which was loaded into WEKA. In order to obtain better result, feature selection was used for selecting the attributes to be used for the classification. In this research, cross validation method was used.

Random forests, Support Vector Machine (SVM) (LibSVM), Naïve bayes and K-Nearest Neighbor (KNN)(IBK) algorithms were used for this study.

### Data Description

The dataset contains medical records of 299 heart failure patients that was collected at the Faisalabad Institute of Cardiology and at the Allied Hospital in Faisalabad (Punjab,Pakistan) in 2015 [3,4]. The patients consisted of 105 women and 194 men with the age bracket of 40 to 95 years. The dataset contains 13 features, which present clinical, body, and lifestyle information as shown in Table 1.

Table1: Meanings, measurement units, and intervals of each feature of the dataset derived from [3]

S/N	Feature	Explanation	Measurement	Range
1	Age	Age of the patient	Years	[40,...95]
2	Anemia	Decrease of red blood cells or haemoglobin	Boolean	0,1
3	High blood pressure	If a patient has hypertension	Boolean	0,1
4	Creatinine phosphokinase (CPK)	Level of the CPK enzyme in the blood	mcg/L	[23,...,7861]
5	Diabetes	If the patient is diabetic	Boolean	0,1
6	Ejection fraction	The percentage of blood that leaves that heart after each contraction	Percentage	[14,...,80]
7	Sex	Male or Female	Binary	0,1
8	Platelets	Platelets in the blood	kiloplatelets/mL	[25.01,...,850.00]
9	Serum creatinine	Level of creatinine in the blood	mg/dL	[0.50,...,9.40]
10	Serum sodium	Level of sodium in the blood	mEq/L	[114,...,148]
11	Smoking	If the patient is a smoker	Boolean	0,1
12	Time	Follow-up period	Days	[4,...,285]
13	Death Event (target))	If the patient died during the follow-up period	Boolean	0,1

### Data Pre-processing

Based on the dataset collected, 6 of the attributes are Boolean including the target class (Anemia, High blood pressure, Diabetes, Sex, Smoking and death event) were converted into two categories, in order to make the dataset usable for classification task. Diabetes, smoking, anemia and high blood pressure 0 and 1 attributes were converted to false and true respectively, while death event 0 and 1 were converted to alive and died respectively also sex 0 and 1 were converted to female and male respectively. In this process it was discovered that the dataset was highly skewed (imbalanced), a novel hybrid of oversampling and

undersampling method called SMOTE+ENN [18] was used to alleviate the class imbalance problem.

### Data Selection

The data selection phase involves understanding the datasets and selecting the attributes which will produce the necessary data needed to infer the knowledge been sought for was then carried out. This is also known as feature selection which is a process of identifying the subset of data from large dimension of data [19]. Time which is the period of follow-up was removed from the dataset which was also done by [3,4].

Attributes contributing more to the development of the model were derived using SymmetricalUncertAttributeEvaluator (SU), InfoGainAttributeEvaluator (IG), GainRatioAttributeEvaluator (GR), CorrelationAttributeEvaluator (CO) coupled with ranker search method. Table 2 presents a summary of the attributes and how the algorithms ranked them.

Table 2: Summary of evaluators' ranking of each attribute of the dataset

Feature	SU	IG	GR	CO
Serum creatinine	0.22054 3	0.30073 4	0.17228 9	0.33559 6
Ejection fraction	0.19676 7	0.30616 4	0.14372 3	0.30851 6
Smoking	0.07610 5	0.06618 1	0.08736 8	0.28818 9
Age	0.07151	0.06842 4	0.07341 8	0.28883 2
Serum sodium	0.07050 4	0.06863 4	0.07110 4	0.20655 9
Diabetes	0.03608 7	0.03424 2	0.03737 9	0.21455 2
High blood pressure	0.02298 8	0.03424 2	0.02321	0.17621 8
Sex	0.02010 2	0.0187	0.02127 8	0.15908 7
Anemia	0.00033 5	0.00032 8	0.00033 6	0.02133
Creatinine phosphokinase	0	0	0	0.00095 7
Platelets	0	0	0	0.06639 6

The first four highest ranked attributes by the evaluators as best influencing heart failure survival are: Serum creatinine, Ejection fraction, smoking, and age. Hence, they are selected for the classification problem.

**Classification**

After the data preprocessing, Random forest, Naïve bayes, Support vector machines (SVM) and K-nearest neighbors algorithms were implemented using Waikato Environment for Knowledge Analysis (WEKA). It is a tested and trusted open source software for machine learning developed at the University of Waikato, New Zealand [20]. Cross validation was selected as the test mode option with 10 as the number of folds and death event attribute was set as the target to be predicted for the classification. This process was done 5 times coupled with changing the random seed starting from 1 -5 for the process for validation purposes.

**RESULTS AND DISCUSSION**

In this section, we first describe the results we obtained for the survival prediction on the complete dataset before data selection

followed by the results we obtained after data selection. Then we evaluated the results obtained to what [3] obtained.

The algorithms were implemented as stated in the previous section. The performance measures which includes Matthews correlation coefficient (MCC) [21], Receiver Operating Characteristic Area Under Curve (ROC AUC), Precision-Recall Area Under curve (PR AUC) which is best used to determine the accuracy of imbalanced dataset [22] and F-Measure which are gotten from the confusion matrix which is used to determine how well a classification has performed [23] by reporting the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Table 3 and Table 4 shown below give the details of the average performance of the classification based on MCC, F-Measure, accuracy, PR AUC and AOC RUC after the process was repeated five times coupled changing of random seed.

Table 3: Details of performance measure of the classification before feature selection

Metho d	MCC	F-Measure	Accuracy (%)	ROC RUC	PR AUC
Random forests	0.542 4	0.7772	77.74834	0.770 6	0.716 6
KNN	0.487 6	0.7504	75.09934	0.742 6	0.696 8
Naïve Bayes	0.485 4	0.7456	75.2318	0.815 6	0.783 6

Table 4: Details of performance measure of the classification after feature selection

Metho d	MCC	F-Measure	Accuracy (%)	ROC AUC	PR AUC
Random forests	0.654 6	0.8316	83.1788	0.906	0.904 6
KNN	0.440 4	0.7266	72.649	0.722	0.678
SVM	0.411	0.7124	71.72186	0.698 8	0.650 6

To this end, we review recent research on the heart failure dataset, as described in Table 5. The results of accuracy, MCC, F-Measure, ROC AUC and PR AUC criteria for the models were obtained according to the 10-fold cross validation method compared to previous studies.

The parameters such as MCC considers the values in the confusion matrix (TP, TN, FP, and FN) and the higher the MCC value (closer to 1) implies the better the classifier performs and both classes are predicted, likewise for F-Measure, ROC AUC and PR AUC values also.

Table 5. The performed works for heart failure survival dataset with the 10-fold cross validation method.

Reference	Methods	No. Subset Selection	Features	MCC	F-Measure	Accuracy	ROC AUC	PR AUC
[3]	Random forests	2		+0.418	0.754	0.585	0.698	0.541
In our study	Random forests	2		+0.575	0.792	0.7947	0.886	0.876
[3]	Random forests	11		+0.384	0.547	0.740	0.800	0.657
In our study	Random forests	11		+0.5424	0.7772	0.7775	0.7706	0.7166
[3]	Naïve Bayes	11		+0.224	0.346	0.696	0.589	0.437
In our study	Naïve Bayes	11		+0.4854	0.7456	0.7523	0.8156	0.7836
[3]	KNN	11		-0.025	0.148	0.624	0.493	0.323
In our study	KNN	11		+0.4876	0.7504	0.7509	0.7426	0.6968

Taking a look at Table 5, it can be seen that the proposed method outperforms other methods in terms of accuracy, MCC, F-Measure, ROC AUC, and PR AUC. It implies that the 4 features extracted are the most informative ones about heart failure survival.

### Discussion

Our results not only show that better accuracy can be obtained from well-handled imbalanced dataset but also, more accurate result could be obtained by feature selection. This aspect is encouraging for healthcare industry should many laboratory test results and clinical features were missing from the electronic health record of a patient; health workers could still be able to predict patient survival by just analysing the ejection fraction, serum creatinine, age and smoking values.

Based on our analysis also, some interesting results that differ from the original dataset curators study [4], Ahmad and colleagues, affirmed that smoking status has nothing to do with heart failure survival of patient. Instead, smoking status is a major determinant after serum creatinine and injection fraction.

We advise that additional confirmatory studies need to be carried out before this machine learning procedure can be taken up into clinical practice.

### CONCLUSIONS

Conclusively, it can be said that serum creatinine, injection fraction, smoking status and age are the major determinants in predicting heart failure survival. Also in this work we have shown the necessity of dealing with imbalanced dataset for classification.

In addition, this study does not have any data on the blood group, genotype and physical future such as the weight, height and body mass index. Therefore, it would be interesting to include it in a future study.

### REFERENCES

- [1] O.N. Emuoyibofarhe, S. Adebayo, A. Ibitoye, O.A. Madamidola and T. Aderibigbe, "Predictive System for Heart Disease Using a Machine Learning Trained Model," *International Journal of Computer (IJC)*, Vol. 34, No. 1, pp 140-152, 2019.
- [2] National Heart Lung and Blood Institute (NHLBI). Heart failure., [Online] Available: <https://www.nhlbi.nih.gov/health-topics/heart-failure> Accessed 2020-04-10

- [3] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fractional one", *BMC Medical Informatics and Decision Making*, Vol. 20, No. 16, 2019
- [4] T. Ahmad, A. Munir, S. H. Bhatti, M. Aflab, M.A Raza, "Survival analysis of heart failure patients: A case study", vol. 12, no. 7, 2017, DOI: 10.1371/journal.pone.0181001
- [5] J. H. Joloudari et al, "Coronary Artery Disease Diagnosis; Ranking the Significant Features Using a Random Trees Model" *International Journal of Environmental Research and Public Health*, vol. 17, No 731,2020, DOI: 10.3390/ijerph17030731
- [6] WHO. Fact sheet on CVDs. *Global Hearts*. World Health Organization. 2016.
- [7] S. E. Awan, M. Bennamoun, F. Sohel, F.M. Sanfilippo, and G. Dwivedi, "Machine learning based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics". *ESC heart failure*, Vol.6, No.2, pp 428-435, 2019.
- [8] C. M. Bishop., "Pattern Recognition and Machine Learning". Springer ISBN 978-0-387-31073-2, 2006.
- [9] J. H. Joloudari, H. Saadatfar, A. Dehzangi, S. Shamshirband, "Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection". *Inform. Med. Unlocked* 17, 100255, 2017.
- [10] G. Ahmad, M. A. Khan, S. Abbas, A. Athar, B. S. Khan, M. S. Aslam, "Automated Diagnosis of Hepatitis B Using Multilayer Mamdani Fuzzy Inference System". *J. Healthc. Eng.* 2019, 6361318, 2019.
- [11] Y. Wang, L. Kung, S. Gupta, S. Ozdemir, "Leveraging big data analytics to improve quality of care in healthcare organizations: A configurational perspective". *Br. J. Manag.* Vol.30, pp 362–388, 2019.
- [12] D. H. Smith, E. S. Johnson, M. L. Thorp, X. Yang, A. Petrik, R. W. Platt, and I. K. Crispel. "Predicting poor outcomes in heart failure." *Permanente J.* Vol.15 No.4, pp4–11, 2011.
- [13] F. M.Zahid, S. Ramzan, S. Faisal, I. Hussain. "Gender based survival prediction models for heart failure patients: a case study in Pakistan." *PLoS ONE*. Vol. 14, No. 2, 2019.

- [14] J. Cai, J. Luo, S. Wang, and S. Yang. "Feature selection in machine learning: A new perspective." *Neurocomputing* Vol. 300, pp. 70–79, 2018.
- [15] S. Mojriani, G. Pinter, J. Hassannataj Joloudari, I. Felde, N. Nabipour, L. Nadai, A. Mosavi, "Hybrid Machine Learning Model of Extreme Learning Machine Radial Basis Function for Breast Cancer Detection and Diagnosis: A Multilayer Fuzzy Expert System." *Preprints* 2019, doi:10.20944/preprints201910.0349.v1
- [16] S. Liu, M. Motani, "Feature Selection Based on Unique Relevant Information for Health Data." *arXiv* 2018, arXiv:1812.00415.
- [17] UCI, "Machine Learning Repository," [Online] Available: <https://archive.ics.uci.edu/ml/index.php> Accessed on 2020-04-02
- [18] S. O. Folorunsho and A. B. Adeyemo, "Alleviating classification problem of imbalanced dataset", *African Journal of Computing and ICT*, Vol. 6, No. 1., pp. 137-144, 2013.
- [19] J. Crisóstomo, et al. "Hyperresistinemia and metabolic dysregulation: the close crosstalk in obese breast cancer." *Endocrine*, vol. 53, no. 2, pp. 433-442, 2016.
- [20] Weka, Available at: [www.cs.waikato.ac.nz/ml/weka](http://www.cs.waikato.ac.nz/ml/weka) Accessed on 2020-04-03
- [21] B. W. Matthews. "Comparison of the predicted and observed secondary structure of T4 phage lysozyme." *Biochim Biophys Acta (BBA)-Protein Struct.* Vol. 405 No. 2, pp. 442–51, 1975.
- [22] T. Saito and M. Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets". *PLoS ONE*. Vol. 10, No. 3, 0118432, 2015.
- [23] P. Diez, "Smart Wheelchairs and Brain-Computer Interfaces," *ScienceDirect*, 2018