Migration Letters

Volume: 21, No: S11 (2024), pp. 670-678

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Advancing Heart Failure Outcomes Using Machine Learning: A Data Science Breakthrough

[1]Haseeb Ur Rehman, [2]Maria Malik, [3]Muhammad Awais, [4]Saira Sharif, [5]Farhatul-ain

Abstract

The scientific community and media are paying significant attention to machine learning. Such algorithms offer a lot of possibilities to improve patient care and personalize it in the medical field, including heart failure diagnosis and treatment. This study examined the risk variables in heart failure patients. In 2015, 299 heart failure patients were collected for our analysis. Using the Decision Tree, K-Nearest Neighbor, Support Vector, and Random Forest classifier, a heart failure outcome model is created with an accuracy of 80%, 77%, 82%, and 88% respectively. Current uses of machine learning techniques to diagnose, classify, and predict heart failure are evaluated in an overview of machine learning geared toward practicing clinicians.

1 Introduction

For centuries, Drops; were the disease entity that we today call heart failure (Ntusi & Mayosi, 2009 Keulenaer & Brutsaert, 2011; Damman & Testani, 2015). The first documented observations on heart failure were provided by Egyptians as bloodletting to discharge fluid accretion and shortage of breath (Yernault, 2004; Choffnes, 2016; Choffnes, 2016;[1] Snowden, 2019). Cases of dyspnea, liquid retention, and weakness compatible with heart failure were described by Greeks (Moore & Porter, 2008; Zhang, 2012). Roman used the flowering plant "Drimia Maritima" for the treatment of dropsy, which contains cardiac glycosides (Aswal, 2019; Manganyi et al., 2021). China and ancient India are also known for pertaining descriptions of heart failure (Savarese & Lund, 2017; Jie et al., 2017; Lee & Fan, 2022). Heart failure is the final stage of all heart disorders and is a significant source of morbidity and mortality. (Davis, 2000; Kemp &Conte; 2012; Bhatti, 2016). Heart failure terminology is not the same as heart attack, in which muscles of the heart die due to clotting in arteries supplying the heart, or cardiac capture (de Silva, 2013; Heart disease. ABC-CLIO. Patel, 2017; Cenko, 2021).

Heart disease is a leading reason of death, accumulating one-third deaths of in the whole world in 2019 (Roth, 2020). The largest number of deaths occurred due to heart failure in China, and the second highest number of deaths accounted for in India, followed by Russia, the United States, and Indonesia (Forouzanfar, 2017; Roth et al., 2020). The lowest heart failure rates were observed in Japan, France, and Peru (Mackay, 2004; Finegold et al., 2013; Bots, 2017). From 1990 to 2019, heart failure cases doubled from 271 million to 523 million in 2019, and the

[1,4,5]University of Management and Technology.
[2]Comsats University Islamabad.
[3]Govt Graduate College, Sheikhpura
Corresponding Author E.mail id: mariamalikkhan@gmail.com

death rate rose to 18.6 million from 12.1 million. Among these 18.6 million 9.6 million were men and 8.6 were women (Roth, 2020; Baddour, 2020). According to the WHO Ranking, Pakistan is the thirtieth place in the world. Deaths reached 240,720 or 16.49% of the total deaths. In Pakistan, males are more victims of heart failure than females (Ishaq, 2021). Diabetes, high blood pressure, sweet beverages, and obesity are all factors contributing to heart attacks in Pakistan (Misra et al., 2017; Raza et al., 2021). Some reviews related to classification methods applied for the prediction of heart failure outcomes are as follows: Davie et al. (1997) evaluated the signs of heart failure in patients of Western General Hospital, UK. Open access to echocardiography (ECO) service was set up to detect left ventricular systolic dysfunction in patients consecutively 259 patients were studied and 41 patients with left ventricular systolic function studied significantly in ECO. Patients with clinical features not needed ECO, and were allocated without such diagnostic features. Davis et al. (2000) studied the prevalence of heart disease in United Kingdom. They estimated 5% of admissions in medical wards of hospitals, with over 100,000 annual admissions in hospitals. They found 3 to 20 per 1000 people and even this exceeds 100 per 1000 patients over the age 65 years. The annual number of heart failures was about 1-5 per 1000, which comparatively doubles for each decade of life after 45 years of age. They examined that with the passage of time, overall survival in patients with weakened cardiac function will increase due to advancement in the management of acute myocardial infarction. According to their study, it was difficult to diagnose heart failure clinically in the early times, but due to advancements in modern drug treatment heart failure risk was reduced as it slowed the rate of disease progression. Davide Chicco and Giuseppe Jurman (2020) analyzed a dataset of 299 patients with heart failure in the year 2015. They applied different machine-learning techniques to predict both patient survival and features associated with significant risk factors. According to their findings, both serum creatinine and ejection fraction were the most significant features in predicting the survival of heart failure patients from medical records.

Sajjad et al. (2017) analyzed the survival of heart failure patients hospitalized at the "Institute of Cardiology and Allied Hospital" in Faisalabad, Pakistan, over nine months. The mortality model was created using Cox regression, which took into account age, serum creatinine, serum sodium, ejection fraction, anemia, platelets, creatinine phosphokinase, blood pressure, diabetes, gender, and smoking as probable causes of mortality. They used the Kaplan-Meier plot to demonstrate the high intensity of mortality. Blood pressure, ejection fraction, renal failure, age, and anemia were all risk factors for death in cardiac patients. Meng F. et al. (2019) conducted the first retrospective multicenter study in China. They discovered that sudden cardiac death was a highly important component in heart failure patients with poor left ventricular ejection fraction, despite its lack of therapeutic relevance for a variety of reasons.

2 Materials and Methods

2.1 Data Collection

The UCI machine learning repository dataset is used to obtain the Heart failure named Heart failure clinical records data set. There are 299 cases (194 male and 105 female). The data was split into two sets: training and testing, with a 70:30 ratio. The test data set has 89 samples, whereas the training data set contains 209 samples. The dataset has 13 features that report clinical, physical, and lifestyle data, which are briefly described below (table 1). Certain qualities are binary: High anemia Factors to consider include blood pressure, diabetes, gender, and smoking. The 10-folder cross-validation method is utilized.

Table 1 **Description of Features**

	Attributes	**Description**
I	Age	Age of the patient (years)
II	Anaemia	Decrease of red blood cells or hemoglobin (boolean)
III	High Blood Pressure	If the patient has hypertension (boolean)
IV	Creatinine phosphokinase (CPK)	Level of the CPK enzyme in the blood (mcg/L)
V	Diabetes	If the patient has diabetes (boolean)
VI	Ejection fraction	Percentage of blood leaving the heart at each contraction (percentage)
VII	Platelets	Platelets in the blood (kiloplatelets/mL)
VIII	Gender	Male and female (binary)
IX	Serum creatinine	Level of serum creatinine in the blood (mg/dL)
X	Serum sodium	Serum sodium: level of serum sodium in the blood (mEq/L)
XI	Smoking	If the patient smokes or not (boolean)
XII	Follow-up	Period (days)
XIII	Death event	If the patient deceased during the follow-up period (boolean)

2.2 The Performance Measure Indices

A few performance indices are used to evaluate how well machine learning algorithms perform. The confusion matrix is a table that displays how many outcomes from the tested model were correctly and incorrectly predicted. The mentioned matrix's features are: Many adversely projected events that have been positively validated are known as false negatives. True positives are a set of results that have been positively anticipated and verified. False positives are a term used to describe many confidently expected results that are later found to be incorrect. True negatives are several adversely expected effects that turn out to be unfavorable. To estimate the parameters, TP, FP, TN, and FN are combined to form a confusion matrix of the actual and estimated class.

The performance of the proposed system is formulated as:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$\text{Sensitivity} = \frac{(TP)}{(TP + FN)}$$

$$\text{Specificity} = \frac{(TN)}{(TN + FP)}$$

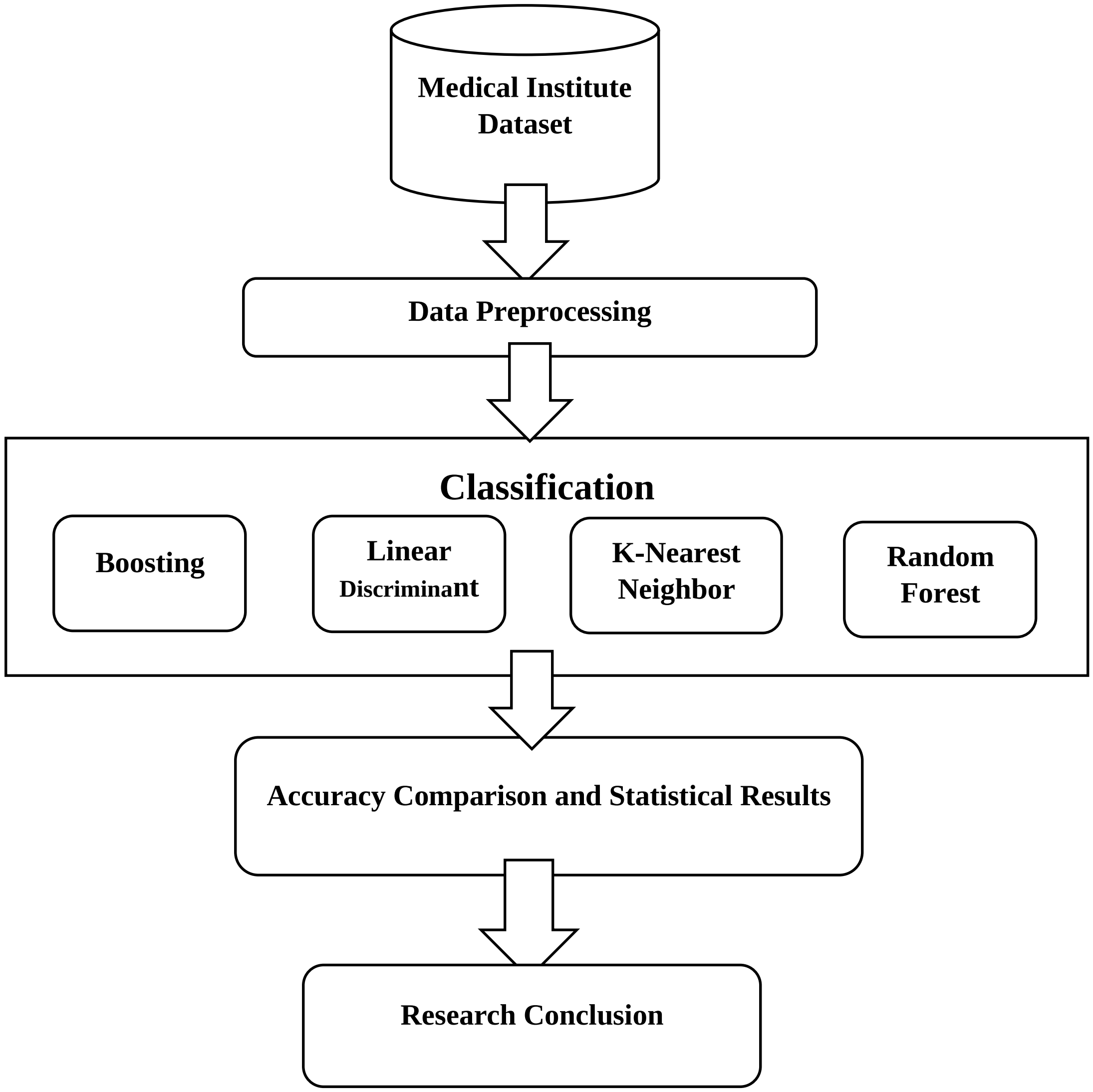


Figure 1

2.3 Decision Tree Classifier

A decision tree is a non-parametric supervised learning technique used for classification and regression tasks. It has a hierarchical tree structure that includes a root node, branches, internal nodes, and leaf nodes.

2.4 K-Nearest Neighbors (KNN)

KNN is a supervised machine learning algorithm that can be used for regression as well as classification. KNN categorizes newly entered data based on the degree of similarity between the new and old classes. The Kth neighbor is closest to the unknown data. To boost prediction accuracy, the model should be trained on a large dataset, and the value of k in the k-nearest neighbor approach must be an odd integer.

2.5 Support Vector Machine

The support vector machine (SVM) is a common classification approach that works well with nonlinear datasets. Pisner and Schnyer explained this grouping. The support vector machine linear model is used for mental health prediction studies. Diseases, sentiment analysis, and other such issues.

2.6 Random Forest

Random forest tree is a machine learning algorithm that is commonly used for behavior analysis and prediction purposes. An enormous number of separate decision trees are combined to represent a specific occurrence in the classification of the input data. Each decision tree casts a vote indicating the expected data class whereas a random decision tree takes into account many decision tree instances independently and produces a prediction output based on the votes of the majority. An optimal attribute from the whole set of input attributes is randomly selected for each tree, and a decision tree model is created from this attribute by finding the best split using the Gini index. The problem with random forest trees is overfitting, which can be resolved by maximizing the decision trees and leading to high accuracy.

3 Results

Table 2 depicts the background characteristics of the heart failure outcome. In this study 4 ML algorithms were applied to classify the Heart failure outcomes. Performance parameters (Accuracy, Sensitivity, Specificity, and AUC values) were used to compare the predictive performance of the proposed algorithms. We see that the test data accuracy of the Random Forest is 88% which means that the algorithm is 88 percent correct for the prediction.

Table 2 Performance Indicators of all Four Machine Learning to Predict Heart Failure Outcome

	Decision Tree	K-Nearest Neighbor	Support Vector	Random Forest
Support	89	89	89	89
Accuracy (%)	0.809	0.775	0.82	0.888
Precision	0.811	0.763	0.816	0.885
Recall	0.809	0.775	0.82	0.888
False Positive Rate	0.213	0.317	0.237	0.169
False Discovery Rate	0.218	0.265	0.196	0.126
F_Score	0.81	0.762	0.814	0.884
Area Under Curve (AUC)	0.787	0.808	0.763	0.908

ROC Curves Plot for Decision Tree

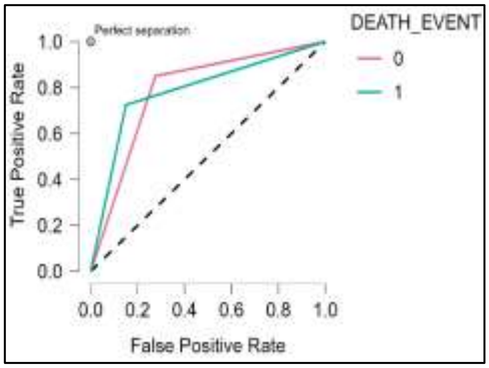


Figure 2

ROC Curves pot for K-NN

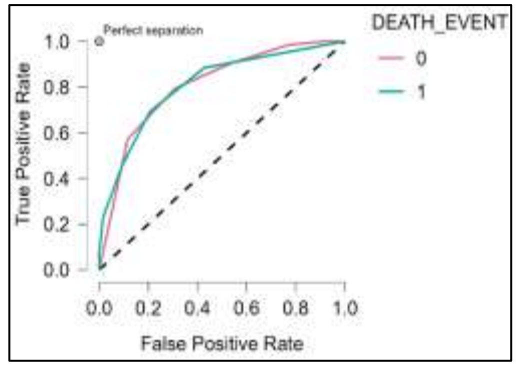


Figure 3

ROC Curves Plot for Support Vector

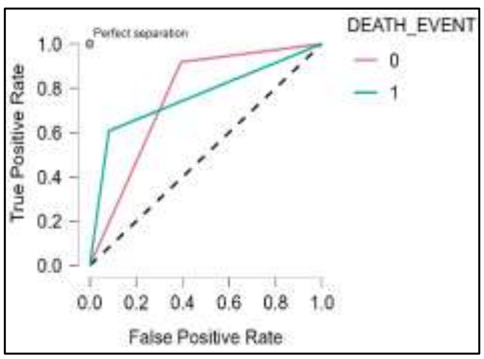


Figure 4

ROC Curves Plot for Random Forest

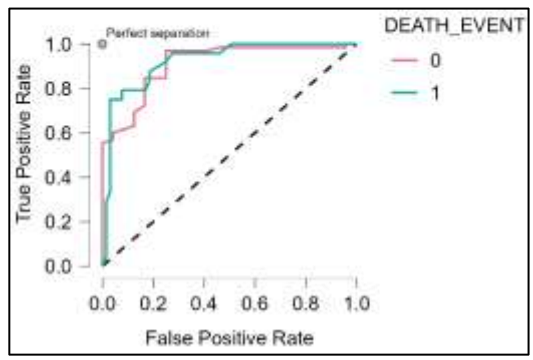


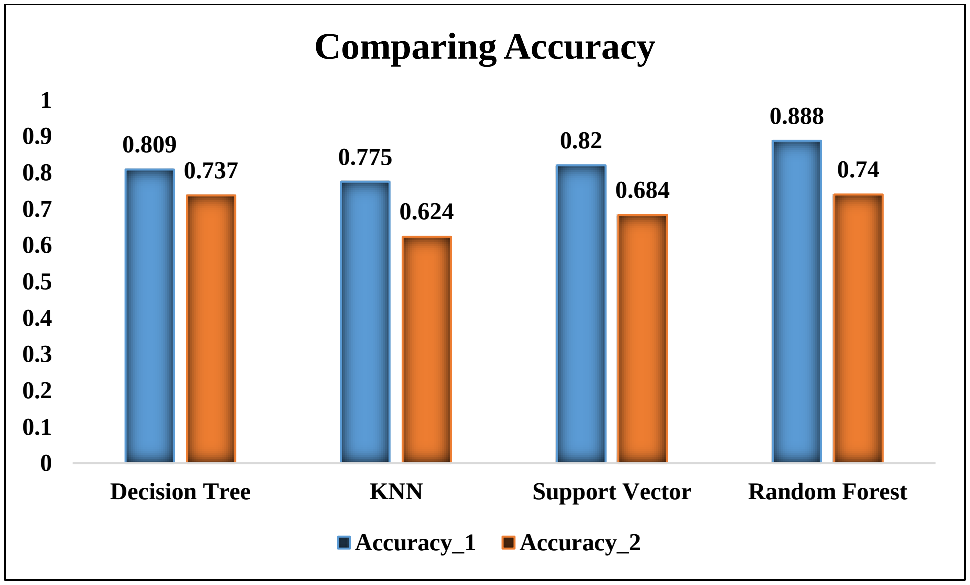
Figure 5

4 Comparative Study

This study comprises the relative comparison of the proposed classifier with the classifier in the literature.

5 Conclusion

It is essential to create a system that can forecast heart diseases precisely and effectively given the rise in fatalities caused by heart diseases. The goal of the study was to identify the most effective machine-learning algorithm for heart disease identification. Using data from the UCI machine learning repository, this study examines the accuracy scores of the Decision Tree, K-Nearest Neighbor, and Support Vector and Random Forest algorithms for predicting heart disease. According to the study's findings, the Random Forest algorithm is the most effective algorithm for predicting heart disease, with an accuracy score of 88% as compared to the one utilized in this investigation, which will help to deliver better results and aid medical practitioners in accurately and successfully predicting cardiac disease.



References

Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M., & Raza, M. A. (2017). Survival analysis of heart failure patients: A case study. PloS one, 12(7), e0181001.

Aswal, S., Kumar, A., Semwal, R. B., Chauhan, A., Kumar, A., Lehmann, J., & Semwal, D. K. (2019). Drimia indica: A plant used in traditional medicine and its potential for clinical uses. Medicina, 55(6), 255.

Ateba, C. N. (2021). Bulbous Plants Drimia: "A Thin Line between Poisonous and Healing Compounds" with Biological Activities. Pharmaceutics, 13(9), 1385.

Baddour, L. M., Addolorato, G., Ammirati, E., Mensah, G., Johnson, C., Rwegerera, G. M., & Roth, G. A. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: update from the GBD 2019 Study.

Bhatti, N. K., Karimi Galougahi, K., Paz, Y., Nazif, T., Moses, J. W., Leon, M. B., & Ali, Z. A. (2016). Diagnosis and management of cardiovascular disease in advanced and end-stage renal disease. Journal of the American Heart Association, 5(8), e003648

Bots, S. H., Peters, S. A., & Woodward, M. (2017). Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of aging between 1980 and 2010. BMJ global health, 2(2), e000298.

Cenko, E., Badimon, L., Bugiardini, R., Claeys, M. J., De Luca, G., de Wit, C., & Tousoulis, D. (2021). Cardiovascular disease and COVID-19: a consensus paper from the ESC Working Group on Coronary Pathophysiology & Microcirculation, ESC Working Group on Thrombosis, and the Association for Acute CardioVascular Care (ACVC), in collaboration with the European Heart Rhythm Association (EHRA). Cardiovascular research, 117(14), 2705-2729.

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC medical informatics and decision making, 20(1), 1-16.

Choffnes, D. (2016). Nature's pharmacopeia: A world of medicinal plants. Columbia University Press. Choffnes, D. (2016). Concepts of Ethnomedicine. In Nature's Pharmacopeia (pp. 2-25). Columbia University Press.

Damman, K., & Testani, J. M. (2015). The kidney in heart failure: an update. European heart journal, 36(23), 1437-1444.

Davie, A. P., Francis, C. M., Caruana, L., Sutherland, G. R., & McMurray, J. J. (1997). Assessing diagnosis in heart failure: which features are any use? QJM: monthly journal of the Association of Physicians, 90(5), 335-339.

Davis, R. C., Hobbs, F. D. R., & Lip, G. Y. H. (2000). ABC of heart failure: history and epidemiology. BMJ: British Medical Journal, 320(7226), 39.

de Silva, R. A. (2013). Heart disease. ABC-CLIO.

De Keulenaer, G. W., & Brutsaert, D. L. (2011). Systolic and diastolic heart failure are overlapping phenotypes within the heart failure spectrum. Circulation, 123(18), 1996-2005.

Finegold, J. A., Asaria, P., & Francis, D. P. (2013). Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. International journal of cardiology, 168(2), 934-945.

Forouzanfar, M. H., Liu, P., Roth, G. A., Ng, M., Biryukov, S., Marczak, L., & Murray, C. J. (2017). Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. Jama, 317(2), 165-182.

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE access, 9, 39707-39716.

Jie, Z., Xia, H., Zhong, S. L., Feng, Q., Li, S., Liang, S., & Kristiansen, K. (2017). The gut microbiome in atherosclerotic cardiovascular disease. Nature communications, 8(1), 1-12.

Kemp, C. D., & Conte, J. V. (2012). The pathophysiology of heart failure. Cardiovascular Pathology, 21(5), 365-371.

Mackay, J., Mensah, G. A., & Greenlund, K. (2004). The atlas of heart disease and stroke. World Health Organization.

Manganyi, M. C., Tlatsana, G. S., Mokoroane, G. T., Senna, K. P., Mohaswa, J. F., Ntsayagae, K., ... &

Savarese, G., & Lund, L. H. (2017). Global public health burden of heart failure. Cardiac failure review, 3(1), 7.

Lee, A. P. W., & Fan, Y. (2022). Battling Congestion in HFpEF: Can Echocardiography Be Our Crystal Ball? JACC: Asia, 2(1), 85-86.

Meng, F., Zhang, Z., Hou, X., Qian, Z., Wang, Y., Chen, Y., & Zou, J. (2019). Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retroprospective multicentre registry in China. BMJ open, 9(5), e023724.

Misra, A., Tandon, N., Ebrahim, S., Sattar, N., Alam, D., Shrivastava, U., & Jafar, T. H. (2017). Diabetes, cardiovascular disease, and chronic kidney disease in South Asia: current status and future directions. bmj, 357.

Moore, J. P., & Porter, S. (2008). Cardiac disease. Tidy's Physiotherapy, 212.

Ntusi, N. B., & Mayosi, B. M. (2009). Epidemiology of heart failure in sub-Saharan Africa. Expert review of cardiovascular therapy, 7(2), 169-180.

Patel, M. R., Calhoon, J. H., Dehmer, G. J., Grantham, J. A., Maddox, T. M., Maron, D. J., & Smith, P. K. (2017). ACC/AATS/AHA/ASE/ASNC/SCAI/SCCT/STS 2017 appropriate use criteria for coronary revascularization in patients with stable ischemic heart disease: a report of the American College of Cardiology appropriate use criteria task force, American Association for Thoracic Surgery, American Heart Association, American Society of Echocardiography, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society of Thoracic .... Journal of the American College of Cardiology, 69(17), 2212-2241.

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., ... & GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. Journal of the American College of Cardiology, 76(25), 2982-3021.

Roth, G. A., Mensah, G. A., & Fuster, V. (2020). The global burden of cardiovascular diseases and risks: a compass for global action. Journal of the American College of Cardiology, 76(25), 2980-2981.

Raza, Q., Nicolaou, M., Cay, F., & Seidell, J. (2021). Association of dietary intake and dietary habits with risk of cardiovascular disease among immigrant Pakistanis living in the Netherlands. Journal of the Pakistan Medical Association, 71(1), 219-227.

Snowden, F. M. (2019). Epidemics and society: From the black death to the present. Yale University Press.

Yernault, J. C. (2004). The history of pleural disease. LUNG BIOLOGY IN HEALTH AND DISEASE, 186, 1-22.

Zhang, K. M. (2012). Impact of Goal Compatibility on Self-Care Adherence Among Patients with Congestive Heart Failure.