Heart Failure Prediction

Corey D. Cooke [1] **Harshvardhan** [1] **Pragya Kandel** [1] **Vanessa Lama** [1]

Abstract

Heart failure affects around 26 million people worldwide, according to Centers for Disease Control and Prevention (2020). The project aims to work on a dataset of 299 patients to predict heart failure and validate the prediction model based on several characteristics available with through the dataset. We will implement various stages of machine learning techniques such as feature selection and classification to develop this predictive model.

1. Introduction

Heart failure is defined as a clinical condition when the heart is unable to pump enough blood and oxygen to support other organs in the body (Centers for Disease Control and Prevention, 2020). It affects around 26 million people worldwide making it a global pandemic (Savarese & Lund, 2017). In United States alone, 6.2 million adults suffer from heart failure and was the cause of 3,39,800 deaths in 2018 (Savarese & Lund, 2017). This situation is likely going to get worse and is predicted to affect 8 million people by 2030 (Lam et al., 2011). Heart failure not only affects the health of people but also impacts the economy of the country (United States of America). In 2012, when there were 5.7 million people with heart failure, it cost the nation an estimated 30.7 billion dollars, which was about 10 percent of total health expenditure (Centers for Disease Control and Prevention, 2020). The cost is estimated to increase by 128 percent between 2012 and 2030.

Despite heart failure prevention being one of the top priorities for physicians, the forecasting of heart failure related events has not reached a high accuracy (Al Aref et al., 2019; Dunn et al., 2007). With the current advances in computing systems, machine learning can be an effective tool for predicting the survival of patients with heart failure symptoms and for the detection of risk factors that can indicate impending heart failure (Dunn et al., 2007). With the available data on electronic heath records, machine learning can be used in understanding the patterns and correlations amongst a patient's data (Awan et al., 2019b). This will give insight on the causes, severity, and understanding of the risk factors.

In this project, we would like to study the given data in the light of various machine learning tools. As part of our course, we were introduced to theory behind and practical implementation of feature selection, pre-processing of dataset, classification, neural networks, and support vector machines. The goal of this project is to use the theory and tools we've learned from the course to implement heart failure prediction and compare the performance of our model with respect to existing models and their performances. We will use performance metrics such as accuracy, run time, and confusion matrices to interpret our validation results.

2. Dataset Details

The dataset that will be extracted and used for the project is a reduced version of the complete dataset which was collected between April and December of 2015. There are 299 patients in the study. The information from patients were collected at the Faisalabad Institute of Cardiology and Allied Hospital in Faisalabad, located in the Punjab province of Pakistan. The dataset contains 13 features reporting on bodily, clinical, and lifestyle information about the patients. Some features such as anaemia, past instance of high blood pressure, diabetes, patient's sex, and smoking habits are binary in nature, while the rest are continuous. (For more details on dataset features see Chicco & Jurman (2020).)

The dataset was originally used in a study by Ahmad et al. (2017) and was later reused by Chicco & Jurman (2020), which is the paper cited on Kaggle, the site hosting this dataset [1].

3. Survey of Previous Works

Ahmad et al. (2017) used various factors contributing to heart failure such as age, ejection fraction, anemia, and blood pressure to predict mortality using the Cox regression. The result was further validated by computing calibration

[1]University of Tennessee, Knoxville. Correspondence to: Corey D. Cooke <ccooke7@vols.utk.edu>.

COSC 522 Final Project.

[1]Larxel. (2019). Heart Failure Prediction: 12 Clinical Features For Predicting Death Events. *Kaggle.* Accessed at https://www.kaggle.com/andrewmvd/heart-failure-clinical-data on November 11, 2021.

slope and discrimination ability via bootstrapping and found that age, renal dysfunction, blood pressure, ejection factor, and anemia significantly affected mortality. The prediction of survival of patients with heart failure of various features contributing to heart failure was done using various machine learning approaches such as linear regression, tree-based approach, artificial neural network (ANN), support vector machine (SVM), k-nearest neighbors (kNN) and Naïve Bayes (NB).

The use of machine learning tools predicted that the two features of serum creatinine level and ejection fraction were enough to predict the survival of heart failure patients. It predicted the survival with higher accuracy than the use of original dataset (Chicco & Jurman, 2020).

In addition to studies on various factors contributing to heart failure, there have also been studies to compare the various machine learning algorithm on prediction of heart failure. The study by Gürfidan & Ersoy (2021) compared various machine learning algorithms for classification of death related to heart failure. They found that classification accuracy was highest with SVM (0.83) and lowest in kNN and decision tree classification (0.73).

Blecker et al. (2016) compared heart failure identification using electronic health record data through different types of algorithms. The first one, heart failure on problem list, second, presence of 1-2 characteristics, third logistic regression of 30 clinically relevant structured data elements, fourth, machine learning approach using unstructured notes and lastly, machine learning approach using structured and unstructured data. Upon analysis, they found that the area under ROC was highest in fifth approach (0.974) and lowest in third approach (0.953). While comparison of positive prediction value it was highest in the first algorithm (0.96) and lowest in the third approach (0.68).

Awan et al. (2019a) used the 47 features available in this dataset to predict the situation and outcomes of the patients. Building on their past study (Awan et al., 2019b) that explored the ability of machine learning models to predict heart failure re-admissions and deaths, they check the importance of different features in the study. The previous study was complex and simpler models with similar accuracies were possible (Awan et al., 2019a). Note that the outcome of interest is whether an individual was readmitted to hospital or died, i.e. a binary variable.

Awan et al. (2019a) uses several methods to select the important features. Initially, they do Fisher's t-test (Fisher, 1992) to determine which variables were powerful enough to distinguish between the possible outcomes. The variables that could distinguish were included in the study. Furthermore, they also used Chi-squared test (Pearson, 1900) to determine the variables that had significant association with the outcome variable.

They also try several other feature-selection methods like sequential forward selection and sequential backward selection which selects variable that improves accuracy most and reduces accuracy least, respectively (see Marcano-Cedeño et al. (2010) and Papatheocharous et al. (2012) for details of the method). Finally, they used Principal Components Analysis (Wold et al., 1987) for selection of variables. Notably, they did not use Fisher's Linear Discriminant Analysis (Cohen et al., 2014) for the purpose. The highest accuracy achieved was by the use of following variables: age, type of index admission, visit to an allied health professional in the last 6 months, length of hospital stay, use of antineoplastic and immunomodulating agents in the last 6 months, and history of HF, chronic kidney disease and depression (Awan et al., 2019a). It presented an AUC of 0.62 — same as the model which used all 47 variables.

4. Technical Approaches

In this section, we present a brief overview of ten different machine learning models used in this project.

4.1. Maximum Posterior Probability Classifiers

Maximum Posterior Probability (MPP) methods come from the set of statistical learning methods. The predicted class $\hat{\omega}$ is the one that maximizes the posterior probability, i.e.:

$$\begin{aligned}\hat{\omega} &= \arg\max_{\omega_j} P(\omega_j|x) \\ &= \arg\max_{\omega_j} \frac{p(x|\omega_j)P(\omega_j)}{p(x)},\end{aligned} \tag{1}$$

where $p(x) = \sum_{i=1}^{c} p(x|\omega_j)P(\omega_j)$. In general, it is assumed that the likelihood $p(x|\omega_j)$ is distributed normally with a mean $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$. The choice of covariance structure decides the type of MPP classifier: Euclidean, Mahalanobis or Quadratic. The classification decision is made using the cost function, i.e. the observation belongs to the class which has the highest MPP.

Euclidean Classifier In this classifier, the covariance for both classes is assumed to be:

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \sigma^2\mathbf{I} \tag{2}$$

This produces a linear decision boundary. For more details, see Bishop (2006).

Mahalanobis Classifier In this classifier, the variance of the features is considered to be different from each other

but considered same across classes. That is, if $\mathbf{\Sigma}_1$ is the sample covariance matrix computed from class 1 training samples and $\mathbf{\Sigma}_2$ is the sample covariance matrix computed from class 2 training samples, then the average of the two, i.e.

$$\mathbf{\Sigma} = \frac{\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2}{2} \tag{3}$$

is used for this classifier while calculating the MPP. For more details, see Bishop (2006).

Quadratic Classifier In this classifier, $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are computed separately for each class and used in (1). This results is a quadratic decision boundary.

4.2. k-Nearest Neighbours

The k-nearest neighbors (kNN) classifier chooses the test sample that is the closest to the majority vote among the k-nearest neighbors in Euclidean space. Ties are broken randomly. To calculate the distance between points, Euclidean distance is used:

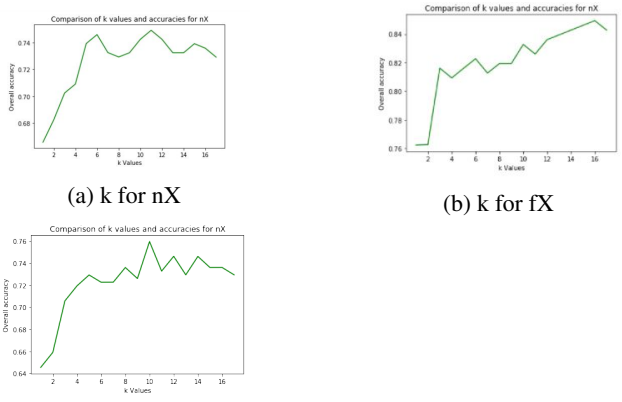
$$d(x_1, x_2) = \sqrt{(x_1 - x_2)'(x_1 - x_2)}. \tag{4}$$

Finally, the observation x_i gets assigned to the class based on majority voting.

$$P(Y = j|X = x) = \frac{1}{K} \sum_{i \in A} I(y^{(i)} = j), \tag{5}$$

where $I(.)$ is the binary indicator function giving value one when $y^{(i)} = j$ and zero otherwise.

For the implementation of kNN to our dataset, we used several values of k and tested the overall accuracy to pick the best k value for each dataset variation (nX, fX, pX). We can see this comparison in the the figures 1



(a) k for nX

(b) k for fX

(c) k for pX

Figure 1. Comparison of k values for kNN classifier.

4.3. Back-propagation Neural Networks

Back-propagation neural networks (BPNN), for the purposes of this document, can be described by a composition of affine functions with a sigmoid activation function. The sigmoid function $\sigma(x)$ is given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{6}$$

If the scalar argument x is replaced by a vector argument $\mathbf{x}$, the sigmoid is applied component-wise. We can define a single layer i of the neural network as:

$$f_i(\mathbf{x}) = \sigma(\mathbf{W}_i\mathbf{x} + \mathbf{b}_i) \tag{7}$$

where $\mathbf{W}_i$ is a weight matrix and $\mathbf{b}_i$ is a bias vector. The total neural network output $\mathbf{y}$ is given by the composition:

$$\mathbf{y} = f_L(f_{L-1}(\ldots(f_2(f_1(\mathbf{x})))\ldots)) \tag{8}$$

where the inner $L-1$ layers are referred to as "hidden layers" (Goodfellow et al., 2016). Training is performed by using gradient descent and recursively propagating derivatives back to the input layer, hence the name.

4.4. Support Vector Machines

Support vector machines (SVM) are the combination of a linear method with a kernel trick applied to the input to increase the dimensionality of the feature space. The objective function to be maximized is the margin between the support vectors. Quadratic programming (QP) is used to solve this optimization problem (Bishop, 2006). For more details on the method, see Noble (2006).

4.5. Random Forest

The random forest (RF) method is a combination of a decision tree with bootstrap aggregation ("bagging"). A "forest" of trees are trained using bagging as a training and fusing method. The general idea of a bagging method is that a combination of learning models will increase the overall result. A major advantage of random forest method is that it works for both classification and regression problems.

The algorithms was originally designed by Breiman (1996). In the original version, a_n observations are drawn at random from the original dataset. Then, a tree-based model is built using these data points. At each node, a split is performed so as to maximise CART-criterion, or the cost function. The process is repeated several times with different set of data points. Finally, all such trees are merged together to form random forest. For more details on the method, see Biau & Scornet (2016).

4.6. Orthogonal Matching Pursuit

The orthogonal matching pursuit (OMP) method uses least squares to find the optimum linear model subject to a spar-

sity constraint (number of nonzero coefficients) scikit-learn (2021).

4.7. Ridge Regression

Ridge regression is a penalised least square method that uses an L_2 norm as a regularisation parameter. It penalises for model complexity and thus performs an indirect variable selection. The method forces coefficients of terms that least improve over a simple linear model to be close to zero, but not exactly equal to zero. For more details on the method, see Hastie et al. (2017).

4.8. Logistic Regression

Logistic regression is a linear model which has uses log odds of an event as the response. A common transformation to convert the class labels into posterior probabilities is called *logit* transformation.

$$P(\omega_i|X = x) = \frac{\exp(\beta_0 + \beta' x)}{1 + \exp(\beta_0 + \beta' x)}. \tag{9}$$

The decision boundary is constructed using the set of points for which the *log-odds* are zero. This hyperplane is defined as the set $\{x : \beta_0 + \beta' x = 0\}$. For more details, see Hastie et al. (2017).

4.9. Passive Aggressive Classifier

The passive aggressive classifier is a linear model that is similar to the perceptron, but with the inclusion of a regularization parameter. Training is performed in a online method (one sample at a time); if the model output is correct, then no adjustments are made ("passive"), otherwise an adjustment is made ("aggressive"). See Crammer et al. (2006) and scikit-learn (2021) for more details.

4.10. Fusion Techniques

We also implemented fusion using two classifiers, the squared Mahalanobis Distance classifier and the Quadratic Classifier (cases 2 and case 3 from our machine learning course. Then, we performed two Bayesian-based approach to fusion, Naive Bayes combination and Behavior-knowledge space (BKS).

5. Exploratory Data Analysis

Before jumping to feature selection and classification, we deemed it appropriate to visualize the dataset to get some idea of how it works. A bar plot showing patient outcomes is shown in Figure 2. It can be seen that approximately one-third of the patients in our dataset died of heart failure (class 1). A histogram of each feature is shown in Figure 3. It is noteworthy that some features are distributed in normally while some others have a bimodal distribution. Some features are presented as binary variables. Many statistical methodologies assume that the input dataset is distributed normally and thus we need to be cognisance of this potential violation.

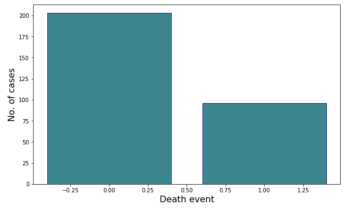


Figure 2. Barplot showing patient outcomes. 1 represents death and 0 represents survival.

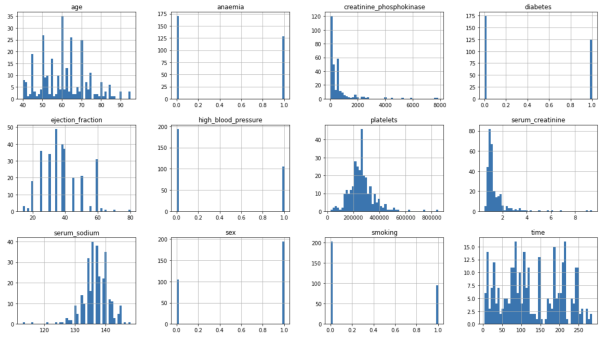


Figure 3. Histogram of 12 features present in the dataset. These 12 features are used to determine patient outcomes.

6. Machine Learning Pipeline

In this project, we are trying are implementing a machine learning approach to predict heart failure from a selected set of clinical and biometric indicators. To perform this task, we will be applying the machine learning (ML) techniques which we learned in class. The data we have would be first preprocessed, followed by classification, and lastly, post-processing. The ML approach we will implement in this project is shown below in Figure 4.

Design of Experiment In a natural experiment or randomized clinical trial, we would conduct a factorial design for each input feature. Then, via experimentation we would conclude about each potential combination (Montgomery, 2017).

Our design is quasi-experimental because we do not have a control group to compare against for effectiveness of our treatment variables. Our quasi-experimental design had 12 clinical features determining patient outcome. The goal is to determine patient outcome with the highest accuracy.

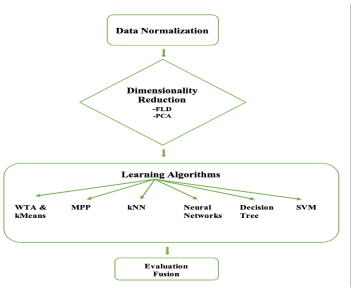


Figure 4. Schematic diagram the for machine learning pipeline of our project.

7. Preprocessing of Data

We are using the heart failure prediction data from Kaggle for this project. [2] The data has 13 categories and 299 samples. The information is divided into two classes based on death events: class 1 and class 0.

In the first step, data would be standardized by subtracting the mean and dividing by standard deviation. We denote to this variation of our dataset as nX. Then, the standardized data is further pre-processed using the Fisher's linear discriminant (FLD) method and principal component analysis (PCA) for dimensionality reduction as described later.

Dimensionality Reduction We applied the two dimensionality reduction methods on our raw data, which resulted in two more processed datasets as follows:

1. FLD dataset, denoted as `fX`,

2. PCA dataset with $< 5\%$ error rate (i.e. keep $\geq 95\%$ of the variance), denoted as `pX`.

In this report, we randomly split the data for validation using n-fold cross validation method with $n = 5$.

7.1. Overall Accuracies

For the list of classifiers and fusion methods, we supplied them with the nX, fX, and pX datasets. Below is the list of all the overall accuracies. Tables 1 to 3 report the accuracies from different methods. These tables allows us to pick the highest overall accuracy for different types of pre-processed dataset applied to different classification and fusion approaches.

[2]Larxel. (2019). Heart Failure Prediction: 12 Clinical Features For Predicting Death Events. *Kaggle.* Accessed at https://www.kaggle.com/andrewmvd/heart-failure-clinical-data on November 11, 2021.

Dataset with standardization (nX)	
Case 1	0.79
Case 2	0.81
Case 3	0.75
kNN (k = 6)	0.74
BPNN	0.90
SVM	0.76
Random Forest	0.78
Passive Aggressive Classifier	0.74
BKS (case 1 + case 2)	0.81
NB (case 1 + case 2)	0.83

Table 1. Overall accuracies for the nX dataset with classification and fusion approaches.

Dataset with FLD (fX)	
Case 1	0.82
Case 2	0.84
Case 3	0.84
kNN (k = 13)	0.83
BPNN	0.83
SVM	0.81
Random Forest	0.85
Passive Aggressive Classifier	0.82
BKS (case 1 + case 2)	0.84
NB (case 1 + case 2)	0.84

Table 2. Overall accuracies for the fX dataset with classification and fusion approaches.

Dataset with PCA (pX)	
Case 1	0.83
Case 2	0.82
Case 3	0.74
kNN (k = 10)	0.75
BPNN	0.83
SVM	0.76
Random Forest	0.69
Passive Aggressive Classifier	0.77
BKS (case 1 + case 2)	0.82
NB (case 1 + case 2)	0.83

Table 3. Overall accuracies for the pX dataset with classification and fusion approaches.

8. Best Performance

Confusion Matrices The confusion matrices are reported in Table 4 to 6. The best performing classifiers were BPNN for nX dataset, Random Forest (only slightly) for the fX dataset, and BPNN and Naive Bayes fusion gave similar accuracies for pX dataset. Below are the confusion matrix

for each case reflecting these outcomes.

	Predicted Positive	Predicted Negative
Actual Positive	199	4
Actual Negative	15	81

Table 4. Confusion matrix for complete dataset with all features (nX).

	Predicted Positive	Predicted Negative
Actual Positive	188	15
Actual Negative	28	68

Table 5. Confusion matrix for data with features selected using FLD (fX).

	Predicted Positive	Predicted Negative
Actual Positive	191	12
Actual Negative	21	75

Table 6. Confusion matrix for data with features selected using PCA w/ 5 percent error rate restriction (pX).

ROC Curves The ROC curves is presented in Figure 5. BPNN didn't produce any ROC curve as it's hyper-parameter indirectly controlling the prior probability.

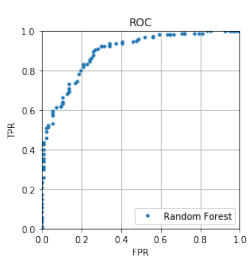
Comparison with Literature Review We also compared the results from our algorithm implementations with the implementations from different research papers. Accuracies for kNN, SVM and NB Fusion are taken from Gürfidan & Ersoy (2021). Accuracies for BPNN, Random Forest and Passive-Aggressive classifier are taken from Chicco & Jurman (2020). The comparison is presented in Figure 6.

We could not find any literature that had implemented MPP classifiers or the BKS classifier on this data. Our best method, BPNN, performed significantly better than the state-of-the-art method. We tried finding the hyperparameters used by Chicco & Jurman (2020) but due to incomplete information we couldn't replicate the same. We hypothesise that this drastic improvement is likely because of the choice of initial states for our model or theirs, choice of hyper-parameters, or a combination of both reasons.

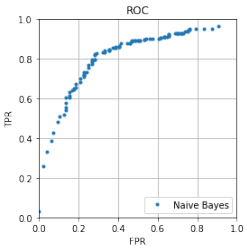
9. Concluding Remarks

Heart failure is a cause of concern in modern world. Our goal was to develop a machine to predict patient outcomes given other details. In our replication study we tried many classifiers on cardiac arrest and heart failure dataset. We found Back-propagation Neural Networks to be the best performing one with an accuracy of 91%.

The study is limited in its power. The data points we have are limited. Heart failure isn't a trivial issue and any model based on only 299 samples might lead to over-fit no matter what train/test/validate proportion we choose. Furthermore, the features are not well ascribed. It is unclear why certain features are treated as binary while others are continuous, even when the medical community would consider them continuous. There were a few past studies like Awan et al. (2019a) that considered manual non-algorithmic feature selection methods like t-test, Chi-squared test and F-test. These selection methods can be an extension to our project's current state. However, we didn't include it in our study as it was beyond the scope of this course and paper.



(a) fX with Random Forest



(b) pX with Naive Bayes Fusion

Figure 5. ROC curves for best performance classification techniques for each variation of our dataset. BPNN didn't produce any ROC curve as its hyper-parameter indirectly controls the prior probability.

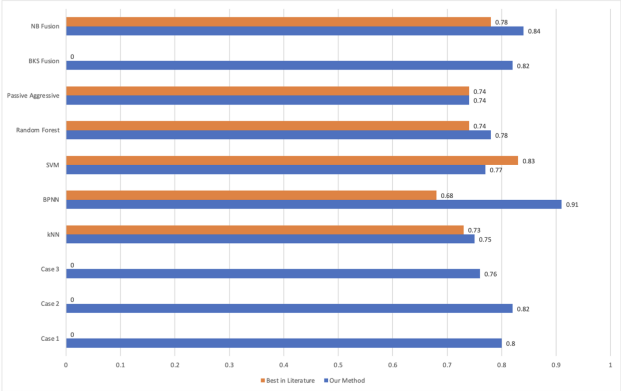


Figure 6. Comparison with state-of-art implementation from literature review. Accuracies for kNN, SVM and NB Fusion are taken from Gürfidan & Ersoy (2021). Accuracies for BPNN, Random Forest and Passive-Aggressive classifier are taken from Chicco & Jurman (2020). The comparison is presented in Figure 6.

Awan et al. (2019a)

Acknowledgement

We present our sincere thanks to Prof Hairong Qi, Senjuti Dutta and Taher Naderi for their kind support and help throughout the course and this project.

Task Allocation

The tentative task allocation between the group members has been described below. However, this was only considered as guide and not as a strict rule. All of us helped each other improve the team's contribution. This project wouldn't have been possible without extensive discussions between members on all aspects of the project.

- Corey Cooke: Implementing classification and fusion methods; code review and re-usability; proof reading report drafts.
- Harshvardhan: Code review and implementation; performance evaluation and reporting; proof reading report drafts.
- Pragya Kandel: Code review and implementation; performance evaluation and reporting; proof reading report drafts.
- Vanessa Lama: Implementing validation and evaluation methods; code review and implementation; proof reading report drafts.

References

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

Al Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., Pandey, M., Maliakal, G., Van Rosendael, A. R., Beecy, A. N., et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal*, 40(24):1975–1986, 2019.

Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F. M., Chow, B. J., and Dwivedi, G. Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death, 06 2019a. URL https://doi.org/10.1371/journal.pone.0218760.

Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F. M., and Dwivedi, G. Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics, 2019b.

Biau, G. and Scornet, E. A random forest guided tour. *Test*, 25(2):197–227, 2016.

Bishop, C. *Pattern Recognition and Machine Learning*. Springer, January 2006. URL https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/.

Blecker, S., Katz, S. D., Horwitz, L. I., Kuperman, G., Park, H., Gold, A., and Sontag, D. Comparison of approaches for heart failure case identification from electronic health record data. *JAMA cardiology*, 1(9):1014–1020, 2016.

Breiman, L. Bagging predictors. *Machine learning*, 24(2): 123–140, 1996.

Centers for Disease Control and Prevention. Heart failure, 2020. URL https://www.cdc.gov/heartdisease/heart_failure.htm.

Chicco, D. and Jurman, G. 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, 2020.

Cohen, P., West, S. G., and Aiken, L. S. *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology press, 2014.

Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y. Online passive aggressive algorithms. 2006.

Dunn, W. B., Broadhurst, D. I., Deepak, S. M., Buch, M. H., McDowell, G., Spasic, I., Ellis, D. I., Brooks, N., Kell, D. B., and Neyses, L. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. *Metabolomics*, 3(4):413–426, 2007.

Fisher, R. A. Statistical methods for research workers. In *Breakthroughs in statistics*, pp. 66–70. Springer, 1992.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.

Gürfidan, R. and Ersoy, M. Classification of death related to heart failure by machine learning algorithms. *Advances in Artificial Intelligence Research*, 1(1):13–18, 2021.

Hastie, T., Friedman, J., and Tisbshirani, R. *The Elements of statistical learning: data mining, inference, and prediction*. Springer, 2017.

Lam, C. S., Donal, E., Kraigher-Krainer, E., and Vasan, R. S. Epidemiology and clinical course of heart failure with preserved ejection fraction. *European journal of heart failure*, 13(1):18–28, 2011.

Marcano-Cedeño, A., Quintanilla-Domínguez, J., Cortina-Januchs, M., and Andina, D. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In *IECON 2010-36th annual conference on IEEE industrial electronics society*, pp. 2845–2850. IEEE, 2010.

Montgomery, D. C. *Design and analysis of experiments*. John wiley & sons, 2017.

Noble, W. S. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

Papatheocharous, E., Papadopoulos, H., and Andreou, A. S. Feature subset selection for software cost modelling and estimation. *arXiv preprint arXiv:1210.1161*, 2012.

Pearson, K. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

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

scikit-learn. Linear models, 2021. URL https://scikit-learn.org/stable/modules/linear_model.html.

Wold, S., Esbensen, K., and Geladi, P. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.