Predictive Tools to Evaluate Cardiovascular Events in Chronic Heart Failure Patients

Maria Carmela Groccia[1][a], Rosita Guido[1][b], Domenico Conforti[1][c] and Angela Sciacqua[2][d]
[1]*Department of Mechanical, Energy and Management Engineering, University of Calabria, Ponte Pietro Bucci 41C, 87036 Rende (Cosenza), Italy*
[2]*Department of Medical and Surgical Sciences, University Magna Graecia of Catanzaro, Italy*

Keywords: Predictive Models, Knowledge Discovery, Machine Learning, Chronic Heart Failure, Decision Tree, Respiratory Rate.

Abstract: In this paper, a Knowledge Discovery task has been implemented with the aim of developing models for predicting cardiovascular worsening events in Chronic Heart Failure (CHF) patients. A set of patients suffering from CHF were enrolled and carefully evaluated through a five-year follow-up. Several predictive models were developed on the collected data and then compared. Among these, the decision tree based predictive model has been analysed by clinical experts. The decision tree is among all the trained and tested models the most simple and interpretable mainly by clinicians because it discovers if-then rules. The extracted rules are compliant with previous clinical studies. Nevertheless, the decision tree achieved lower performance compared to the other predictive models, which conversely to the decision tree are not "clinician friendly" because they do not provide an explanation of the classification decisions.

1 INTRODUCTION

Chronic Heart Failure (CHF) is a complex syndrome caused by the inability of the heart to pump a sufficient amount of blood around the body. Typical symptoms of CHF patients include breathlessness, fatigue, and ankle swelling (McDonagh et al., 2021; Morrissey et al., 2011). Based on the symptoms, the New York Heart Association (NYHA) classification distinguishes the CHF patients in four classes; from NYHA I, without any limitation to physical activity, to NYHA IV, where patients have inability to carry out any type of activity without discomfort (Bredy et al., 2018). CHF confers high risk for cardiovascular (CV) worsening events that cause recurrent hospitalizations and high mortality rate even in patients with mild symptoms (Dunlay et al., 2009). An early prediction of CV worsening events could offer benefits for a preventive treatment, limit serious consequences and improve the quality of care (Ponikowski et al., 2014). Therefore, it could have a relevant advantage on the reduction of hospitalizations and associated costs.

Machine Learning (ML) and Knowledge Discovery (KD) techniques can be applied to allow early prediction of CV worsening events. These methodologies allow to learn knowledge from past experiences through identifying patterns in the data (Fayyad et al., 1996). ML algorithms are able to automatically learn these patterns from past data and apply it to future predictions.

Previous research applied ML and KD techniques in CHF domain for predicting adverse events (Tripoliti et al., 2017; Groccia et al., 2018). In (Tripoliti et al., 2017) a review of several models for predicting the presence of adverse events, such as destabilizations, re-hospitalizations, and mortality is presented. In (Groccia et al., 2018) a temporal weighting approach was applied to risk prediction of major cardiovascular worsening events in CHF patients taking into account the chronology of events.

In this paper, a KD task has been designed and implemented to extract new predictive models that can help clinicians to early detect CV worsening events in patients with CHF. The KD task was conducted on a real dataset collected over a five years follow-up. Several ML algorithms were trained obtaining different

[a] https://orcid.org/0000-0001-7570-8458
[b] https://orcid.org/0000-0003-1744-2166
[c] https://orcid.org/0000-0002-4816-4333
[d] https://orcid.org/0000-0003-1674-5317

Groccia, M., Guido, R., Conforti, D. and Sciacqua, A.
Predictive Tools to Evaluate Cardiovascular Events in Chronic Heart Failure Patients.
DOI: 10.5220/0010829900003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 475-481
ISBN: 978-989-758-552-4; ISSN: 2184-4305

predictive models. Models performance have been evaluated and compared. Decision tree based model has been deeply analysed by clinicians exploiting the possibility to extract simple rules compliant with the previous clinical studies.

2 METHODS

A KD task has been designed and implemented to analyse the collected data and to develop models for predicting CV worsening events in CHF patients. The KD analysis was defined as a predictive task stated as supervised binary classification problem.

Supervised ML approaches learn from a given dataset a function f that predicts an output variable (or class label) y from a feature vector x containing N input variables, such that $y = f(x)$ (Mitchell, 1997; Jo, 2021). In the considered classification problem, the class label is a categorical variable with only two values.

2.1 Collected Data

The collected real dataset contains clinical information of 50 patients with an established diagnosis of CHF and NYHA classes I, II and III. The data was collected at the CHF ambulatory of the Geriatrics Division at the "Mater Domini" University Hospital in Catanzaro, Italy. Patients were followed up every 3 months on an outpatient basis for an average of five years. All patients gave their availability and written consent for participation at the pilot study.

At first outpatient visit, personal data and medical history including date of birth, gender, NYHA class, etiology, cardiovascular history, use of medications, other diseases were collected. At each outpatient visit, vital signs such as Heart Rate (HR), Body Temperature (BT), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Respiratory Rate (RR) and weight were recorded. Dates of specific events (i.e., date of the visits, date at which a CV worsening event occurred) were reported too. During the follow-up, 19 patients presented a CV worsening event. Among these, 8 patients had more than one event. Demographic and clinical characteristics of the patients are summarized in Table 1.

2.2 Preprocessing

In the original format, the dataset is organized in a wide format and consists of 50 rows. Each row contains, in its first columns, the patients data that don't change across time: personal data and medical history. The remaining columns contain the clinical parameters recorder at each visit and the dates of specific events. To perform the classification task, the dataset was converted in a long format. In this format, each patient has data in multiple rows. Each row represents a patient's visit.

The class label was defined based on the occurrence or not of CV worsening events between two consecutive outpatient visits. An instance is a row of the dataset. 31 instances were designated as positive

Table 1: Demographic and clinical characteristics of the patients. NYHA: New York Heart Association; CHF: chronic heart failure; PTCA: percutaneous transluminal coronary angioplasty; ICD: implantable cardioverter defibrillator; TIA: transient ischaemic attack; COPD: chronic obstructive pulmonary disease; CV:cardiovascular.

Characteristic	All patients n=50
Age (years ± SD)	72.5 ± 14.2
Gender	
Male	36 (72%)
Female	14 (28%)
NYHA Class	
I	3 (6%)
II	38 (76%)
III	9 (18%)
CHF etiology	
Ischemic heart disease	23 (46%)
Idiopathic dilatation	9 (18%)
Hypertension	4 (8%)
Valvular diseases	8 (16%)
Valvular diseases + Hypertension	4 (8%)
Alcoholic habit	2 (4%)
Cardiovascular history	
Instable angina	1 (2%)
PTCA	1 (2%)
By-pass	7 (14%)
Atrial flutter	13 (26%)
Pacemaker	3 (6%)
Cardiac resynchronization	1 (2%)
ICD	2 (4%)
Mitral insufficiency	21 (42%)
Aortic insufficiency	4 (8%)
Hypertension	28 (56%)
TIA	2 (4%)
Other diseases	
Diabetes	11 (22%)
Hypothyroidism	1 (2%)
Renal failure	4 (8%)
COPD	5 (10%)
Asthma	1 (2%)
Sleep apnea	4 (8%)
Pulmonary fibrosis	1 (2%)
Gastrointestinal diseases	4 (8%)
Hepatic diseases	3 (6%)
CV worsening events	19 (38%)

instances (patients with CV worsening events) and the remaining 762 instances were designated as negative instances (patients without events).

Input errors were corrected and a new variable that contains the age of patients at admission was created.

The dataset is imbalanced because the number of instances of patients with CV worsening events is much lower than the instances of the patients without events.

With the aim to create predictive models that use few and simple clinical parameters, only the vital signs measured at each outpatient visit were included in the training set. Clinical parameters, i.e., HR, RR, DPB, and SBP are used both to monitoring CHF and as a primary tool regarding patient status.

The dataset was randomly divided into training (70%) and test (30%) set. The training set was used to build the predictive models. The test set instead, was used to evaluate the performance of each model on unseen data. A resampling approach has been adopted to balance the classes in the training set.

2.3 Models Building

Several ML algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Naïve Bayes, Decision Tree and Random Forest were implemented to develop the predictive models.

SVM is a classifier based on statistical learning theory (Cortes and Vapnik, 1995; Burges, 1998). It searches for an optimal hyperplane, in an N-dimensional space, that separates patterns of classes by maximizing the margin. In non-linearly separable dataset, the SVM maps inputs into high-dimensional feature spaces using a kernel function in order to transform it in a linear separable dataset. The most popular kernels used in SVM classification tasks are polynomial kernels and Radial Basis Function (RBF), also called Gaussian kernels.

ANN is a computational model, consisting of a number of artificial neural units called perceptron. They emulate biological neural networks (Krenker et al., 2011). In this work, we used a type of a fully connected, feed-forward artificial neural network named Multilayer Perceptron (MLP). MLP consists of neurons arranged in layers: one input layer, one output layer, and one or more hidden layers.

Naïve Bayes is a probabilistic classification algorithm based on the Bayes Theorem with strong (naïve) independence assumptions between the features (Rish, 2001). This algorithm is based on the assumption that a particular feature in a class is unrelated to the presence of any other feature.

Decision Tree is a non-parametric supervised learning method (Quinlan, 1986). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A Decision Tree consists of nodes and branches. Each node represents an input attribute and a split point on that attribute. The leaf nodes contain an output attribute which is used to make a prediction. Given a new input, the tree is traversed by evaluating the specific input started at the root node of the tree. One of their main advantage is that they are simple to understand and interpret, and they can be visualised.

Random forest (Breiman, 2001) consists of individual decision trees that operate as an ensemble. Each tree is built by applying bagging, which is the general technique of bootstrap aggregating. A simple majority vote of all trees gives the final result. The interpretability of a single decision tree is lost in random forest because many decision trees are aggregated.

The tuning model hyper-parameters has been optimized using a 5-fold cross validation. In k fold cross validation, the dataset is split randomly into k equal sized folds. K iterations are performed and, at each iteration one of the k folds is used as the validation set while all remaining folds are used as the training set. In this process each instance is used for testing exactly once. The resampling was performed only on the folds used as training data as discussed in (Santos et al., 2018).

Table 2: Hyper-parameters tuning. LR: Learning Rate; Mom: Momentum; Ep: No. epochs; CF: Confidence Factor; Iter: No. iterations.

Model	Hyper-Param	Search space	Step	Best value
SVM linear kernel	C	[1,5]	1	1
SVM poly kernel	d	[1,5]	1	3
	C	[1,10]	0.5	10
SVM RBF kernel	γ	[0.01,1.00]	0.01	0.03
	C	[1,10]	0.5	10
ANN	LR	[0.1,1.0]	0.1	0.3
	Mom	[0.1,1.0]	0.1	0.2
	Ep	[400,600]	100	500
Decision tree	CF	[0.10,0.50]	0.05	0.25
Random Forest	Iter	[100,400]	100	100

The hyper-parameters were optimized by searching the best value in a defined range for each ML model. Table 2 reports the hyper-parameters tuning of the tested algorithms. In the third column of the table

there is the range of values defined as search space. The incremented value is denoted in the fourth column as Step. The last column shows the best value. We tested SVM with three kernel functions, i.e., linear kernel, polynomial kernel, and RBF kernel.

Waikato Environment for Knowledge Analysis (WEKA) software, version 3.8.2, was used to build the predictive models by using classification algorithms (Eibe et al., 2016). We used SMO (Sequential Minimal Optimization) algorithm for SVM and J48 for decision tree.

2.4 Models Evaluation

The Area under the ROC curve (AUC), sensitivity, specificity, and Geometric Mean (G-mean) are used to evaluate and compare the predictive performance of the build ML models on the test set.

The confusion matrix was used to define the metrics discussed in this section. Table 3 shows the structure of confusion matrix. Let P and N, be the number of positive and negative instances, respectively. TP and TN, are the number of instances correctly predicted as positive and negative, respectively; FP and FN are the number of instances predicted as positive and negative whereas they belong to the opposite class, respectively.

Table 3: Confusion Matrix for binary classifier.

	predicted positive	predicted negative
actual positive	TP	FN
actual negative	FP	TN

AUC: measures the classifier's ability to avoid false classification. It is the area under the curve of the true positive ratio vs. the false positive ratio and indicates the probability that the model will rank a positive case more highly than a negative case.

Sensitivity: measures the proportion of positive instances that are correctly identified, i.e, it is the ability to predict a CV worsening event. It is defined as

$$\text{Sens} = \frac{TP}{TP+FN}$$

Specificity: measures the proportion of negatives that are correctly identified, i.e, it is the ability to predict patients without CV worsening events. It is defined as

$$\text{Spec} = \frac{TN}{TN+FP}$$

G-mean: takes into account the balance of the classifier's performance on the two classes. It is defined as the geometric mean of sensitivity and specificity as

$$\text{G-Mean} = sqrt(Sens * Spec)$$

3 RESULTS AND DISCUSSION

Table 4 shows the performance of the predictive models on the test set.

Table 4: Results of the predictive models on the test set.

Model	AUC	Spec	Sens	G-mean
SVM linear kernel	0.77	0.79	0.75	0.77
SVM poly kernel	0.70	0.90	0.50	0.67
SVM radial kernel	0.77	0.79	0.75	0.77
Naive Bayes	0.73	0.83	0.50	0.64
ANN	0.81	0.86	0.50	0.66
Random Forest	0.73	0.93	0.12	0.34
Decision Tree	0.57	0.89	0.25	0.47

Following the specific indications of the clinical domain experts' involved in this study, a deep clinical assessment of the decision tree model has been developed since it is easier to understand by clinicians.

The decision tree extracted from the KD task has undergone a post-processing process by the support of clinical domain experts in order to extract few simple rules that could be directly used by clinicians in their daily practice. The tree is reported in Figure 1. The label 0 identifies the absence of risks, while the label 1 identifies risk of CV worsening events.

The three rules constructed by the decision tree to predict CV worsening events in CHF patients are described below.

Rule 1. The first rule suggests a test on RR (root node). The tree identifies the RR as the most relevant parameter for predicting an event. In particular, a CHF patient with RR greater than 20 apm may have an event.

Rule 2. The second rule suggests a new event for patients with a $RR \leq 20$ apm, $HR > 88$ bpm and $DBP \leq 72$ mmHg.

Rule 3. If $RR \leq 20$ apm and $HR \leq 88$ bpm, the third rule suggests reconsidering again these values. If $56 < HR \leq 75$ bpm with a $RR \leq 18$ apm and $DBP \leq 52$ mmHg, a new event may occur with any SBP value.

Regarding the third classification rule, the SHIFT study (Borer et al., 2012) already showed how HR decrease induced by ivabradine led to a decrease in hospitalization rate in patients suffering from CHF with reduced systolic function. Therefore, the model is consistent with currently published data. While the HR cut-off was not foreseen in the study, the classification rule of our model identifies what is the HR

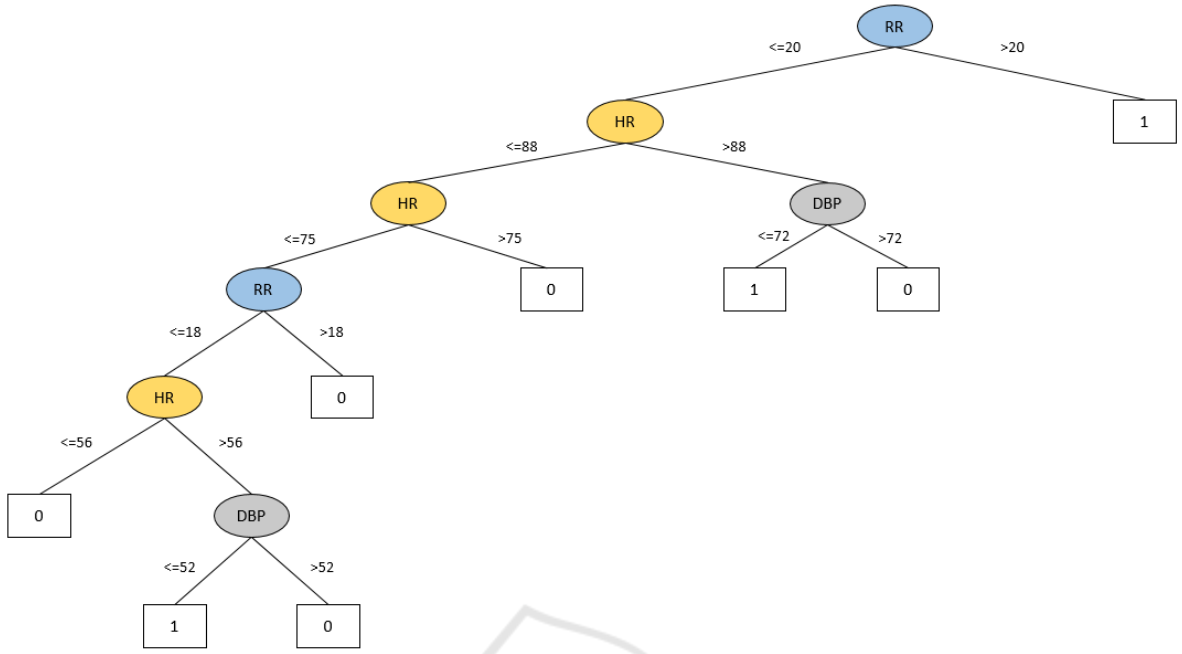

Figure 1: Decision tree.

value associated with an acute destabilization event.

The clinical importance that emerged through this model for the RR is undoubtedly interesting. With regard to the algorithms published in the past, a recent meta-analysis suggests that weight was the most used parameter to monitor CHF patients (Klersy et al., 2009). According to our analysis, weight does not give meaningful contributions to the model and was eliminated from the classification tree. Generally, in this type of model RR is not taken into consideration. Yet RR is an excellent clinical indicator, not only for respiratory system, but also for hemodynamic equilibrium. Indeed, worsening of respiratory diseases negatively affects cardiac function. On the other hand, during the early stages of decompensation lung interstitial congestion can occur, thereby triggering the activation of J receptors which in turn stimulates pulmonary ventilation. It is clinically relevant to take this variable into consideration for the prediction of a worsening event. In this model, RR is considered in more than one classification rule. In particular, the first rule considers exclusively if RR is higher than 20 acts per minute. From a clinical point of view, RR is really useful since tachypnea is an indicator of CV distress, which arises as an attempt of compensatory mechanism that in the long term further promotes decompensation.

Moreover, the second classification rule is based on a group of parameters, including RR, HR and DBP. Of course, all these considerations should not come as a surprise, because clinical parameters can always be affected by compensatory mechanisms. The latter are initially crucial to maintain an adequate cardiac output; however, compensatory mechanisms can contribute to further worsen decompensation. We take into consideration the activation of both sympathetic nervous system and renin-angiotensin system. In addition, we recall the importance of HR as an early clinical indicator of decompensation, closely associated with sympathetic activation and with the consequent positive chronotropic effects expressing an initial compensation mechanism, which however favors a further worsening of the overall clinical status. In fact, HR is the main inducer of myocardial oxygen consumption, which can concur to promote decompensation. These concepts explain why the use of beta-blockers is a cornerstone of CHF treatment.

Finally, it is useful to discuss the role of DBP. Indeed, the second and the third classification rules imply that an event is predicted when the DBP is ≤ 72 mmHg and ≤ 52 mmHg, respectively. Also in this case the system is in agreement with clinical literature. A previous Japanese study (Tsujimoto and Kajio, 2018) had shown that a low DBP value was associated with the onset of CV events and with an increase in the number of hospitalizations due to HF. Notably, in this study DBP values < 70 mmHg were taken into consideration, thereby indicating them as more unfavorable than the $80-89$ mmHg range.

Despite the knowledge extracted with the decision tree is in a form easily to be interpreted and the extracted rules are compliant with previous clinical

studies, this predictive model has lower performance than other models. It has low sensitivity although the reasons behind the low sensitivity could be linked to the low sample size.

The SVM with linear and radial kernel had the best performance on the test set for predicting CV worsening events in CHF patients: $AUC = 0.77$, $specificity = 0.79$, $sensitivity = 0.75$ and $G-mean = 0.77$. This indicate that SVM has a high ability to avoid false classification.

4 CONCLUSIONS

This work presents and compares predictive models based on ML algorithms for the early prediction of CV worsening events in CHF patients using few clinical parameters.

Among the predictive models, SVM with linear and radial kernel had the best performance on the test set. As we showed, the decision tree is among all the trained and tested models the most simple and interpretable mainly by clinicians because it discovers if-then rules as clinicians do. Conversely, although SVM has the best performances, it is not "clinician friendly". The inability of SVMs in providing a simple and understandable interpretation of the classification decisions is one of the main obstacles impeding their application in the clinical practice.

As future work, we will reproduce these experimental models on a larger size study and a shorter interval occurring between two consecutive visits. A follow-up based on shorter intervals could increase both sensitivity and specificity of the models. Another possible application could consider remote monitoring, with the active help of either the patients themselves or their caregivers, to intervene as early as possible to avoid events and subsequent hospitalization. In addition, techniques for rule extraction from SVM could be adopted to ameliorate the aforementioned issue.

ACKNOWLEDGEMENTS

This work has been partially supported by the industrial research and development project "HEARTNET-ICS - Advanced Analytics for Heart Diseases Management" (European Regional Development Fund, Calabria Region Grant J58C17000150006, Italy, 2017-2020)

REFERENCES

Borer, J. S., Böhm, M., Ford, I., Komajda, M., Tavazzi, L., Sendon, J. L., Alings, M., Lopez-de Sa, E., Swedberg, K., and Investigators, S. (2012). Effect of ivabradine on recurrent hospitalization for worsening heart failure in patients with chronic systolic heart failure: the shift study. *European heart journal*, 33(22):2813–2820.

Bredy, C., Ministeri, M., Kempny, A., Alonso-Gonzalez, R., Swan, L., Uebing, A., Diller, G.-P., Gatzoulis, M. A., and Dimopoulos, K. (2018). New york heart association (nyha) classification in adults with congenital heart disease: relation to objective measures of exercise and outcome. *European Heart Journal-Quality of Care and Clinical Outcomes*, 4(1):51–58.

Breiman, L. (2001). Random forests. *Mach Learn*, 45:5–32.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Dunlay, S. M., Redfield, M. M., Weston, S. A., Therneau, T. M., Hall Long, K., Shah, N. D., and Roger, V. L. (2009). Hospitalizations after heart failure diagnosis: a community perspective. *Journal of the American College of Cardiology*, 54(18):1695–1702.

Eibe, F., Hall, M. A., and Witten, I. H. (2016). The weka workbench. online appendix for data mining: practical machine learning tools and techniques. In *Morgan Kaufmann*.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34.

Groccia, M. C., Lofaro, D., Guido, R., Conforti, D., and Sciacqua, A. (2018). Predictive models for risk assessment of worsening events in chronic heart failure patients. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE.

Jo, T. (2021). *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Springer Nature.

Klersy, C., De Silvestri, A., Gabutti, G., Regoli, F., and Auricchio, A. (2009). A meta-analysis of remote monitoring of heart failure patients. *Journal of the American College of Cardiology*, 54(18):1683–1694.

Krenker, A., Bešter, J., and Kos, A. (2011). Introduction to the artificial neural networks. *Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech*, pages 1–18.

McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak, M., Lam, C. S. P., Lyon, A. R., McMurray, J. J. V., Mebazaa, A., Mindham, R., Muneretto, C., Francesco Piepoli, M., Price, S., Rosano, G. M. C., Ruschitzka, F., Kathrine Skibelund,

A., and Group, E. S. D. (2021). 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal*. ehab368.

Mitchell, T. (1997). *Machine Learning*. McGraw Hill.

Morrissey, R. P., Czer, L., and Shah, P. K. (2011). Chronic heart failure. *American Journal of Cardiovascular Drugs*, 11(3):153–171.

Ponikowski, P., Anker, S. D., AlHabib, K. F., Cowie, M. R., Force, T. L., Hu, S., Jaarsma, T., Krum, H., Rastogi, V., Rohde, L. E., et al. (2014). Heart failure: preventing disease and death worldwide. *ESC heart failure*, 1(1):4–25.

Quinlan, J. (1986). Induction of decision trees. *Mach Learn*, 1:81–106.

Rish, I. (2001). An empirical study of the naive bayes classifier. Technical report.

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., and Santos, J. (2018). Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [research frontier]. *ieee ComputatioNal iNtelligeNCe magaziNe*, 13(4):59–76.

Tripoliti, E. E., Papadopoulos, T. G., Karanasiou, G. S., Naka, K. K., and Fotiadis, D. I. (2017). Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. *Computational and structural biotechnology journal*, 15:26–47.

Tsujimoto, T. and Kajio, H. (2018). Low diastolic blood pressure and adverse outcomes in heart failure with preserved ejection fraction. *International journal of cardiology*, 263:69–74.