Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 207 (2022) 1378–1387

Procedia Computer Science

www.elsevier.com/locate/procedia

26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)

Interpretable machine learning models to support differential diagnosis between Ischemic Heart Disease and Dilated Cardiomyopathy

K.Iscra[a,*], A.Miladinović[a,b,c], M.Ajčević[a], S.Starita[a], L.Restivo[d], M.Merlo[d], A.Accardo[a]

[a]*Department Engineering and Architecture, University of Trieste, Trieste, Italy*
[b]*Institute for Maternal and Child Health – IRCCS Burlo Garofolo, Trieste, Italy*
[c]*Science and Research Centre Koper, Institute for Kinesiology Research, Koper, Slovenia*
[d]*Cardiovascular Department, Azienda Sanitaria Universitaria Guiliano Isontina (ASUGI) and University of Trieste, Trieste, Italy*

Abstract

The differential diagnosis between Ischemic Heart Disease (IHD) and Dilated Cardiomyopathy (DCM), particularly in the early stages of the diseases, can often be difficult. Left ventricular ejection fraction (LVEF) and heart rate variability (HRV) analysis are shown to be helpful tools for diagnosing several cardiac diseases. There is a growing interest in application of machine learning techniques to guide the diagnosis. However, often black-box machine learning models create dissatisfaction among clinicians due to the lack of a model interpretability. The aim of our study was to compare the classification performance of interpretable and clinically plausible models applied for early differential diagnosis between DCM and IHD (NYHA $\leq$ 1) based on LVEF and HRV features. The study encompassed 196 IHD and 117 DCM subjects. The models were produced by classification tree, logistic regression and naïve Bayes algorithms considering the set of selected HRV and LVEF features, chosen with the information gain method. The results showed that the most informative features for classification between IHD and DCM were LVEF, LF, NN50, pNN50, and meanRR. The naive Bayes model with classification accuracy of 73.5% outperformed classification tree and logistic regression models with 67.4% and 67.1% accuracies, respectively. We also demonstrated that the produced models together with nomograms allow probabilistic interpretation of the classification output between IHD and DCM, which is an important factor to guide the clinical decision making in differential diagnosis.

* Katerina Iscra. Tel.: +39 0405587130;
E-mail address: katerina.iscra@phd.units.it

Peer-review under responsibility of the scientific committee of the 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)
10.1016/j.procs.2022.09.194

Peer-review under responsibility of the scientific committee of the 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)

Keywords: Computer Aided Diagnosis; Ischemic Heart Disease; Dilated Cardiomyopathy; Interpretable Machine Learning;

1. Introduction

According to the World Health Organization, cardiovascular diseases are the leading cause of mortality worldwide [1]. Ischemic Heart Disease (IHD) and Dilated Cardiomyopathy (DCM) are the most important cardiovascular diseases and the leading cause of mortality worldwide. The incidence of cardiac events is reduced and the prognosis is improved when IHD and DCM is detected early and treated with appropriate medical therapy [2]. DCM, a progressive cardiac muscle disease characterized by left ventricular chamber expansion and contractile failure in the absence of chronic pressure and/or volume overload, is a leading cause of heart failure and the most prevalent indication for heart transplantation globally [3]. IHD, a subtle pathology due to its silent behaviour before developing into unstable angina, myocardial infarction, or possibly, sudden cardiac death, is a major cause of death with a growing clinical impact [4]. The differential diagnosis between IHD and DCM, particularly in the early stages of the diseases, can often be difficult and only invasive, or often not readily available exams can provide a definite diagnosis of IHD and DCM.

The most often used variable for measuring heart function and predicting outcomes in DCM patients is left ventricular ejection fraction (LVEF). However, it has some inherent drawbacks, including late reduction only in patients with advanced cardiac disease, low reliability in patients with left ventricular hypertrophy and volume reduction, very poor inter- and intra-observer variability, and problematic endocardial boundary identification [5]. On the other hand, many authors use heart rate variability (HRV) for the prediction of various heart diseases [6–12]. HRV or the variation over time of the interval between successive heartbeats (RR intervals) is shown to be a helpful tool for diagnosis and it is retrieved from ECG. The HRV measurement is used to quantify cardiac autonomic activity as a result of sympathetic and parasympathetic activity interaction. The pathophysiologic alterations associated with heart diseases, as well as, the resulting changes in HRV, might give crucial prognostic information [13].

There is growing research interest in the development of machine learning models for computer-aided diagnosis that exploits HRV extracted parameters in combination with other available clinical data [14]. Such models may be used to predict illness risk, readmission risk, and the need for treatment, among other things. However, the limited interpretability of produced models still creates dissatisfaction among the clinicians, as they cannot be evaluated on the level of single input features. For this reason, methods such as classification trees, naive Bayes, and linear/logistic regression algorithms are more desirable in medicine [15]. Such interpretable models not only provide output information about a certain disease but also help to intrinsically evaluate the plausibility of the model by examining the selected thresholds and branches in comparison to the existing knowledge reported in the disease diagnostic manuals. Therefore, end-users and healthcare workers can examine the logic behind prediction models in order to accept or reject output [14].

Approaches such as linear/logistic regression, classification trees, and naive Bayes models are employed in several sectors of healthcare, such as toxicology [15,16], endocrinology [18], neurology [16] and cardiology[19–21], due to their high degree of interpretability and ease of use in practice. In particular, Fenny et al.[19] found that the best model for prediction of echocardiographic cardiac resynchronization therapy response was based on the naive Bayes machine learning technique that included nine clinical and demographic variables, such as the New York Heart Association (NYHA) classification, LVEF value, and sex. Salman et al.[20], instead, compared different predictive models of hospital mortality for patients with myocardial infarction. Melillo et al.[21] developed an automatic classifier based on CART for risk assessment in patients suffering from congestive heart failure separating lower-risk patients from higher-risk ones using HRV measures. However, there is a lack of studies that explore HRV parameters to distinguish between IHD and DCM patients.

Nonetheless of the importance of model interpretability, the majority of existing models focus only on accuracy prediction and seldom provide a relevant clinical explanation for their outcomes [14,22]. Interpretability approaches are unquestionably a key issue that must be considered while developing prediction models for healthcare [15].

Hence, the production of clinically plausible machine learning models that can guide diagnosis and can provide information on the feature relevance is desired.

Therefore, we aimed to produce interpretable and clinically plausible models for early differential diagnosis between DCM and IHD based on LVEF and HRV parameters, as well as compare their classification performance.

2. Methods

2.1. Study population and protocol

In this study, we analyzed clinical data and processed ECG signals of 313 subjects. In particular, the study encompassed 196 patients affected by IHD (145 males, aged 70±11, and 51 females, aged 76±9) and 117 patients suffering from DCM (74 males, aged 57±15, and 43 females, aged 65±14). IHD subjects were diagnosed based on clinical and laboratory observations, and coronary angiography was used to confirm the diagnosis [23]. In the three months prior to Holter monitoring, none of the IHD patients had acute coronary syndrome. LVEF, acquired by echographic examination, was obtained by the Simpson biplane method [24]. Patients with DCM were included only if coronary artery disease was not sufficient to explain the dysfunction or if LVEF was less than 50% and cannot be explained by pressure or volume overload [25]. Coronary angiography was conducted on all individuals over the age of 35 who had cardiovascular risk factors and/or had a family history of DCM. Patients with established trigger factors such as toxic insults from alcohol or drugs, as well as tachyarrhythmias, were excluded from the study. Both groups of subjects were receiving beta-blocker medication and were classified as NYHA ≤ 1 according to the New York Heart Association severity scale. The study was performed according to the Declaration of Helsinki and all patients gave written consent.

2.2. Heart Rate Variability acquisition and processing

All of the participants had a 24-hour ECG Holter monitoring session with a three-channel tracking record (Sorin Group, Italy). The ECG signal was sampled at 200Hz and RR intervals were extracted from recordings using SyneScope analysis software (Sorin Group, Italy), which recognizes QRS complexes automatically. The data was analyzed using a MATLAB (MathWorks) program that evaluated the RR segments of 300s. A pre-processing of the RR times series was done to limit the effect of noise and artifacts. The longest ectopic beats sequence or the longest artifact in the studied segments had to be less than 10 seconds long, and the overall duration of artifacts and ectopic beats had to be less than 20% of the segment duration [26]. These segments were interpolated with cubic spline and resampled at 2 Hz, producing the total HRV signal.

Following that, linear and non-linear HRV parameters were obtained for each segment. The linear parameters MeanRR, SDNN, RMSSD, NN50, and pNN50 evaluating RR variability were calculated directly from the RR sequence [27], whereas the absolute powers in the Low (LF=0.04-0.15Hz) and High (HF=0.15-0.40Hz) Frequency bands, related to vagal and sympathetic nerve control on the heart rhythm, were estimated from the interpolated HRV signal in the frequency domain. Furthermore, the latter parameters were used to determine the normalized low and high-frequency powers (LFn, HFn), as well as their ratio (LF/HF). The non-linear analysis was carried out by calculating Poincaré plot parameters (SD1, SD2) reflecting the short and long-term variability [28], and extracting Fractal Dimension (FD) and beta exponent (betaExp) [29] quantifying the complexity of the system generating the signal. Finally, the median of all features from valid 5 min segments during 24h were calculated and used as the input for the classifier. The set of considered linear and non-linear HRV features is summarized in Table 1.

Table 1. The set of linear and non-linear HRV features

HRV parameter	Definition
MeanRR (ms)	mean of RR intervals
SDNN (ms)	standard deviation of RR intervals
RMSSD (ms)	root mean square of the squared differences of successive RR intervals
NN50	number of differences of successive RR intervals greater than 50ms
pNN50	proportion of NN50 divided by the total number of RR intervals
LF (ms^2)	Low Frequency Power (from 0.04 to 0.15Hz)
HF (ms^2)	High Frequency Power (from 0.15 to 0.40Hz)
LF/HF	Low Frequency Power/High Frequency Power
LFn	Low Frequency Power / Total Power
HFn	High Frequency Power /Total Power
betaExp (ms^2/Hz)	Beta exponent
SD1 (ms)	short-term variability of the RR sequence – from Poincarè Plot
SD2 (ms)	long-term variability of the RR sequence – from Poincarè Plot
SD1/SD2	short-term variability / long-term variability of the RR sequence
FD	Fractal dimension

2.3. Feature selection and classification

The models were built considering selected HRV features together with LVEF. The features were chosen based on their correlation with the target parameter, which was computed using the information gain or the expected amount of information. The features that have information gain of at least 0.025 were considered for further modeling.

Classification tree, logistic regression, and naive Bayes methods were employed to produce models capable of differentiating between the two groups (IHD and DCM). These three models were used for diagnostic modeling because of their easy interpretability in the clinical domain [15,22].

The classification tree [30,31] is one of the simples methods and consists of two steps: tree growth and pruning. The tree grows in the first stage by picking from all potential splits those that produce nodes with just one class of components. The result of this stage is the large-tree. Following that, the tree is pruned using a minimal cost-complexity function that takes into account both the number of nodes and the chance of misclassification. The best sub-tree that minimizes the cost-complexity function is the result of this stage. We used Mean Squared Error (MSE) to quantify the impurity of each node during tree growth.

Another method for generating multivariable composites to distinguish two or more groups is to use logistic regression. In general, the sigmoid function argument of a logistic regression classifier can be a linear combination of more than one feature value or explanatory variable. The sigmoid function produces a number between 0 and 1 as its output. The middle value is used as a criterion to determine what belongs in class 1 and what belongs in class 0.

An input that produces a result larger than 0.5 is classified as belonging to class 1. In contrast, if the output is less than 0.5, the matching input is categorized as 0 class [32].

Finally, a naive Bayes classifier is a basic probabilistic classifier that uses the Bayes theorem (from Bayesian statistics) and strong (naive) independence assumptions to classify data [33]. The naive Bayes classifier has the benefit of just requiring a modest quantity of training data to estimate the classification parameters [33]. In reality, the naive Bayes classifier performs admirably, frequently outperforming far more advanced algorithms [34]. Many practical applications of naive Bayes have been demonstrated, including text categorization, medical diagnosis, and system performance management [34].

2.4. Models Interpretability

The selected classifiers were included in this study as they all satisfy the criteria of interpretability. The classification tree modeling, even though in practice it might represent slightly lower accuracy in comparison to the "black-box" models [22], provides better interpretability and practical usability in clinical application. Classification trees and their simple visualization allows clinicians to follow a set of rules and thresholds for selected clinical and HRV parameters. It also helps to evaluate the most important features that are always located at the root of the decision tree.

On the other hand, the logistic regression and naive Bayes can be interpreted by means of nomograms. A nomogram is a basic and self-explanatory visualization that is both helpful and powerful in the diagnostic guidance between two diseases, and it is a graphical representation of numerical relationships. The primary purpose of a nomogram, which was invented by French mathematician Maurice d'Ocagne in 1891, was to allow the user to graphically derive the result of an equation without having to do any complicated calculations. Lubsen and coauthors [35] firstly used nomograms to interpret a logistic regression model. They demonstrated the efficacy of nomograms for predicting an acute myocardial infarction, as they were allowed to be printed on paper and utilized by doctors to determine the likelihood of the outcome. Besides allowing the prediction, the logistic regression and naive Bayesian nomogram reveal the structure of the model and the relative influences of the features on the class probability. In particular, the lengths of the lines correlate to the spans of odds ratios, implying that variables are important. In addition, nomograms allow the calculation of the scores for each feature, and such scores may be utilized not only to obtain the classification outcome but also the probability of having a specific disease [36,37].

2.5. Models' evaluation

The classification accuracy, the area under the curve (AUC), F1 measure, precision, and recall on the dataset were estimated using 5-fold cross-validation. The process was then repeated 5 times, using each of the subsamples only once as the validation data. Therefore, the overall cross-validation performance measures were calculated as a mean of all 5 validation folds. In addition, the confusion matrix and ROC analysis were performed for each model. The approach based on the Bayesian comparison of classification algorithms [38] is used for comparing classification accuracies obtained by cross-validation for each of the selected algorithms. All analysis was carried out in Python Orange3 Data Mining library and toolbox [39].

3. Results

Table 2 shows the selected features by using the information gain as a metric. The features, LVEF, LF, NN50, pNN50 and meanRR with information gain of 0.109, 0.40, 0.037, 0.030, and 0.025 respectively were then used to produce three different classification models (classification tree, logistic regression, and naive Bayes).

Table 2. The set selected features and corresponding information gain

Features	Information Gain
LVEF	0.109
LF	0.040
NN50	0.037
pNN50	0.030
MeanRR	0.025

Table 3 reports the performance of the produced models. In particular, the model produced by naive Bayes showed the highest classification accuracy (73.5%) compared to models produced by logistic regression (67.4%) and classification tree (67.1%). Regarding the AUC, naive Bayes again presented the highest value (0.740), followed by classification tree (0.729) and logistic regression (0.674). Figure 1 depicts the ROC curves for each of the models and class (DCM Figure 1a and IHD Figure 1b). The same trend was also observed for F1, precision and recall and reported in Table 3.

Table 3. Performance measures of the naive Bayes, logistic regression and classification tree models

Model	Accuracy	AUC	F1	Precision	Recall
Classification tree	0.671	0.729	0.669	0.667	0.671
Logistic regression	0.674	0.674	0.636	0.666	0.674
Naive Bayes	0.735	0.740	0.734	0.733	0.735

Confusion matrices obtained for classification tree, logistic regression and naive Bayes models are reported in Figure 2. Confusion matrices show that DCM patients were correctly classified with 53.0% (Figure 2a), 29.1% (Figure 2b) and 62.4% (Figure 2c) with classification tree, logistic regression and naive Bayes, respectively. On the other hand, IHD were correctly classified with 75.5% (Figure 2a), 90.3% (Figure 2b) and 80.1% (Figure 2c) with the same models.

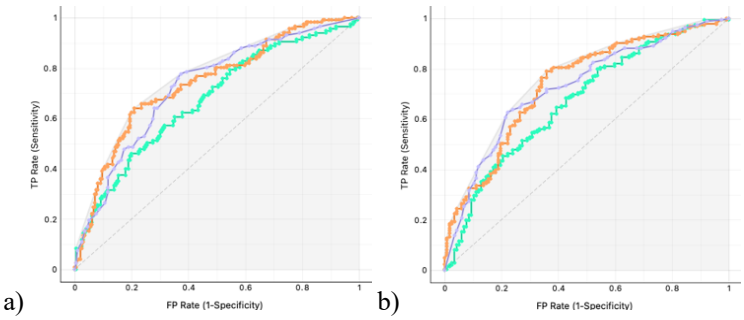

Figure 1. ROC curves for classification tree (purple), logistic regression (green) and naive Bayes (orange) of (a) DCM and (b) IHD patients

a)	DCM	IHD	Σ
DCM	53.0%	47.0%	117
IHD	24.5%	75.5%	196
Σ	53	260	313

b)	DCM	IHD	Σ
DCM	29.1%	70.9%	117
IHD	9.7%	90.3%	196
Σ	53	260	313

c)	DCM	IHD	Σ
DCM	62.4%	37.6%	117
IHD	19.9%	80.1%	196
Σ	112	201	313

Figure 2. Confusion matrices obtained by (a) classification tree; (b) logistic regression; (c) naive Bayes models.

The comparison of cross-validation classification accuracies by means of Bayesian comparison of classification algorithms shows that the probability that classification accuracy (CA) of naive Bayes is higher than CA of the classification tree $p(CA_{naiveBayes}>CA_{tree})$ is 0.923 and classification of logistic regression $p(CA_{naiveBayes} > CA_{logregression})$ is 0.876. The nomogram for naive Bayes approach, that shows the highest classification accuracy, is depicted in Figure 3.

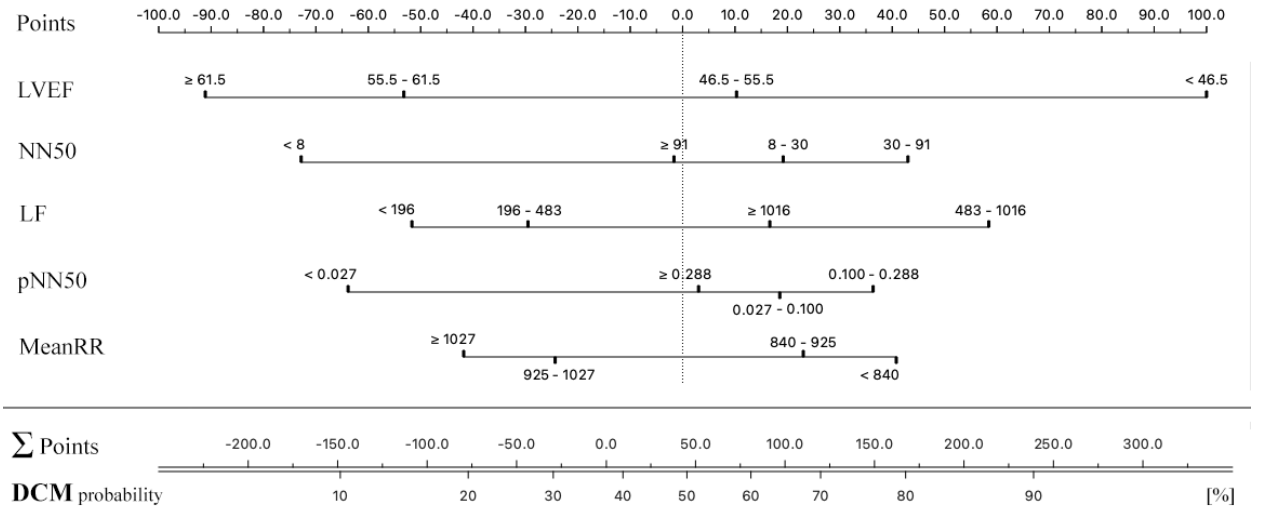

Figure 3. Nomogram for naive Bayes classifier for the DCM patients. The scoring can be obtained as a sum of the score of each individual parameter. For IHD the probability with the same scoring obtained by nomogram can be calculated by subtracting DCM probability from 100.

4. Discussion

Ischemic Heart Disease and Dilated Cardiomyopathy are two of the most important cardiovascular diseases often leading to patient death. Nowadays, these two diseases can only be diagnosed through invasive diagnostic techniques that are expensive and often rarely available. Therefore, there is a growing interest in an auxiliary tool that can guide differential diagnosis between two groups. Several studies [6,40,41] have used various machine learning techniques to differentiate different cardiovascular diseases. However, black-box models often used, provide only final classification output, and they cannot be evaluated on the level of single input features. The evaluation of models by the level of the single features is important to examine the plausibility of the models by comparing the selected thresholds and branches to the existing knowledge reported in the disease diagnostic manual. For this reason, approaches such as classification trees, naive Bayes, and logistic regression not only provide output information about a certain disease but also reveal the structure of the model and the relative influences of the features. Therefore, we aimed at comparing the interpretable and clinically plausible models for early differential diagnosis between DCM and IHD based on LVEF and HRV parameters.

Our results showed that left ventricular ejection fraction, LF, NN50, pNN50, and meanRR, used in naive Bayes approach with classification accuracy (73.5%), AUC (0.740), F1 (0.734), precision (0.733), and recall (0.735), outperforms classification trees and logistic regression.

The feature ranking method selected LVEF as the most important feature, which is in line with the clinical literature as the parameter that can well characterize the DCM group. However, LVEF can be observed in patients with advanced cardiac disease, whereas usually not helpful in the early stages. In addition, LVEF presents a gray zone in the range (40% - 50%) [42] where the diagnosis cannot be performed based only on this parameter. For this reason, other HRV parameters can help toward better classification. In particular, LF is often associated with the joint action of the vagal and sympathetic system [43] that can be altered in both pathologies. In addition, the mean of RR interval (meanRR), number of differences of successive RR intervals greater than 50ms (NN50), the proportion

of NN50 divided by the total number of RR intervals (pNN50) are indicators of the irregularities of the heat rhythms, also characteristic for both diseases.

The selection of interpretable machine learning approaches allows us to further investigate the relationship of these parameters to the output of the classifier. In regard, in the result section, the nomogram of the best approach has been reported. The nomogram reveals the structure of the naive Bayes approach and the influences of the features on the class probability. In particular, the longest length of the feature LVEF implies that is the most discriminatory feature for the two groups. The asymmetry around zero implies that the length to the right side is more important for the classification of DCM and vice versa, the length on the left side implies that the feature is more characteristic of the IHD. For example, NN50 and pNN50 are more characteristic of IHD, which is in line with the literature [44], as they often might present higher heart rate irregularities captured by these two HRV temporal parameters.

Moreover, the nomograms allow further validation of the produced model by observing the thresholds. For example, the clinical literature shows that patients with LVEF ≤ 45% [5] have a high probability of DCM. Indeed, our model depicted in Figure 3 demonstrates the same fact, that the subjects < 46.5 have the highest probability of being classified as DCM if only that parameter is considered. Similar validation can be performed also for other features.

Furthermore, nomograms can be printed on paper and utilized by clinicians to determine the likelihood of being in the DCM or IHD group. The scoring is performed for each of the features and then summed up to obtain the final probability. For example, if the patient has the LVEF value between 46.5 and 55.5, (10 points in the scale), NN50 in the range from 30 to 91 (42 points), LF equal to or higher than 1016 (17 points), pNN50 higher then 0.288 (3 points) and meanRR in the range from 840 to 925 (23 points), and it gets total 95 points, that on the DCM probability score (reported in the nomogram in Figure 3) corresponds to 64%.

In conclusion, we demonstrated that naive Bayes, a simple probabilistic classifier, outperformed other interpretable machine learning approaches (logistic regression and classification tree) with a moderately high classification accuracy of 73.5%. The obtained accuracy is important to guide the diagnostic procedure, as it allows clinicians to better decide the necessity of furthermore complex, invasive exams. We also showed the importance of the interpretability of the machine learning models, as their clinical plausibility can be evaluated by nomograms on the level of the single features. If the model is evaluated by the majority of the features, it can also bring more information about thresholds for yet unknown disease-related HRV parameters. Finally, our study highlighted the importance of nomograms, as a tool which allows a probabilistic classification supporting the clinical decision making in differential diagnosis between early IHD and DCM.

Acknowledgement

Work partially supported by Master in Clinical Engineering, University of Trieste.

References

[1] E. Olvera Lopez, B.D. Ballard, A. Jan, Cardiovascular Disease, in: StatPearls, StatPearls Publishing, Treasure Island (FL), 2022. http://www.ncbi.nlm.nih.gov/books/NBK535419/ (accessed April 1, 2022).

[2] S. Mendis, P. Puska, B. editors Norrving, W.H. Organization, Global atlas on cardiovascular disease prevention and control, World Health Organization, 2011.

[3] B.J. Maron, J.A. Towbin, G. Thiene, C. Antzelevitch, D. Corrado, D. Arnett, A.J. Moss, C.E. Seidman, J.B. Young, American Heart Association, Council on Clinical Cardiology, Heart Failure and Transplantation Committee, Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups, Council on Epidemiology and Prevention, Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention, Circulation. 113 (2006) 1807–1816. https://doi.org/10.1161/CIRCULATIONAHA.106.174287.

[4] A.S. Go, D. Mozaffarian, V.L. Roger, E.J. Benjamin, J.D. Berry, W.B. Borden, D.M. Bravata, S. Dai, E.S. Ford, C.S. Fox, S. Franco, H.J. Fullerton, C. Gillespie, S.M. Hailpern, J.A. Heit, V.J. Howard, M.D. Huffman, B.M. Kissela, S.J. Kittner, D.T. Lackland, J.H. Lichtman, L.D. Lisabeth, D. Magid, G.M. Marcus, A. Marelli, D.B. Matchar, D.K. McGuire, E.R. Mohler, C.S. Moy, M.E. Mussolino, G. Nichol, N.P.

Paynter, P.J. Schreiner, P.D. Sorlie, J. Stein, T.N. Turan, S.S. Virani, N.D. Wong, D. Woo, M.B. Turner, Heart Disease and Stroke Statistics—2013 Update, Circulation. 127 (2013) e6–e245. https://doi.org/10.1161/CIR.0b013e31828124ad.

[5] I.H. Jung, J.H. Park, J.A. Lee, G.S. Kim, H.Y. Lee, Y.S. Byun, B.O. Kim, Left Ventricular Global Longitudinal Strain as a Predictor for Left Ventricular Reverse Remodeling in Dilated Cardiomyopathy, J. Cardiovasc. Imaging. 28 (2020) 137–149. https://doi.org/10.4250/jcvi.2019.0111.

[6] E. Agliari, A. Barra, O.A. Barra, A. Fachechi, L. Franceschi Vento, L. Moretti, Detecting cardiac pathologies via machine learning on heart-rate variability time series and related markers, Sci. Rep. 10 (2020) 8845. https://doi.org/10.1038/s41598-020-64083-4.

[7] L. Zhang, H. Wu, X. Zhang, X. Wei, F. Hou, Y. Ma, Sleep heart rate variability assists the automatic prediction of long-term cardiovascular outcomes, Sleep Med. 67 (2020) 217–224. https://doi.org/10.1016/j.sleep.2019.11.1259.

[8] X. Yan, L. Zhang, J. Li, D. Du, F. Hou, Entropy-Based Measures of Hypnopompic Heart Rate Variability Contribute to the Automatic Prediction of Cardiovascular Events, Entropy. 22 (2020) 241. https://doi.org/10.3390/e22020241.

[9] G. Silveri, M. Merlo, L. Restivo, M. Ajčević, G. Sinagra, A. Accardo, A big - data classification tree for decision support system in the detection of dilated cardiomyopathy using heart rate variability, Procedia Comput. Sci. 176 (2020) 2940–2948. https://doi.org/10.1016/j.procs.2020.09.209.

[10] G. Silveri, M. Merlo, L. Restivo, B. De Paola, A. Miladinović, M. Ajčević, G. Sinagra, A. Accardo, Identification of Ischemic Heart Disease by using machine learning technique based on parameters measuring Heart Rate Variability, in: 2020 28th Eur. Signal Process. Conf. EUSIPCO, 2021: pp. 1309–1312. https://doi.org/10.23919/Eusipco47968.2020.9287800.

[11] A. Accardo, G. Silveri, M. Merlo, L. Restivo, M. Ajčević, G. Sinagra, Detection of subjects with ischemic heart disease by using machine learning technique based on heart rate total variability parameters, Physiol. Meas. (2020). https://doi.org/10.1088/1361-6579/abc321.

[12] A. Accardo, G. Silveri, M. Ajčević, A. Miladinović, L. Pascazio, Influence of smoking and other cardiovascular risk factors on heart rate circadian rhythm in normotensive and hypertensive subjects, PLOS ONE. 16 (2021) e0257660. https://doi.org/10.1371/journal.pone.0257660.

[13] R.E. Kleiger, J.P. Miller, J.T. Bigger, A.J. Moss, Decreased heart rate variability and its association with increased mortality after acute myocardial infarction, Am. J. Cardiol. 59 (1987) 256–262. https://doi.org/10.1016/0002-9149(87)90795-8.

[14] M.A. Ahmad, C. Eckert, A. Teredesai, Interpretable Machine Learning in Healthcare, in: Proc. 2018 ACM Int. Conf. Bioinforma. Comput. Biol. Health Inform., Association for Computing Machinery, New York, NY, USA, 2018: pp. 559–560. https://doi.org/10.1145/3233547.3233667.

[15] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, L. Cilar, Interpretability of machine learning-based prediction models in healthcare, WIREs Data Min. Knowl. Discov. 10 (2020) e1379. https://doi.org/10.1002/widm.1379.

[16] Y. Zhang, Y. Ma, Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia, Comput. Biol. Med. 106 (2019) 33–39. https://doi.org/10.1016/j.compbiomed.2019.01.009.

[17] H. Zhang, J.-X. Ma, C.-T. Liu, J.-X. Ren, L. Ding, Development and evaluation of in silico prediction model for drug-induced respiratory toxicity by using naïve Bayes classifier method, Food Chem. Toxicol. 121 (2018) 593–603. https://doi.org/10.1016/j.fct.2018.09.051.

[18] S. Sossi Alaoui, B. Aksasse, Y. Farhaoui, Data Mining and Machine Learning Approaches and Technologies for Diagnosing Diabetes in Women, in: Y. Farhaoui (Ed.), Big Data Netw. Technol., Springer International Publishing, Cham, 2020: pp. 59–72. https://doi.org/10.1007/978-3-030-23672-4_6.

[19] A.K. Feeny, J. Rickard, D. Patel, S. Toro, K.M. Trulock, C.J. Park, M.A. LaBarbera, N. Varma, M.J. Niebauer, S. Sinha, E.Z. Gorodeski, R.A. Grimm, X. Ji, J. Barnard, A. Madabhushi, D.D. Spragg, M.K. Chung, Machine Learning Prediction of Response to Cardiac Resynchronization Therapy: Improvement Versus Current Guidelines, Circ. Arrhythm. Electrophysiol. 12 (2019) e007316. https://doi.org/10.1161/CIRCEP.119.007316.

[20] I. Salman, Heart attack mortality prediction: an application of machine learning methods, Turk. J. Electr. Eng. Comput. Sci. 27 (2019) 4378–4389.

[21] P. Melillo, N. De Luca, M. Bracale, L. Pecchia, Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability, IEEE J. Biomed. Health Inform. 17 (2013) 727–733. https://doi.org/10.1109/jbhi.2013.2244902.

[22] R. Elshawi, M.H. Al-Mallah, S. Sakr, On the interpretability of machine learning-based model for predicting hypertension, BMC Med. Inform. Decis. Mak. 19 (2019) 146. https://doi.org/10.1186/s12911-019-0874-0.

[23] J. Knuuti, W. Wijns, A. Saraste, D. Capodanno, E. Barbato, C. Funck-Brentano, E. Prescott, R.F. Storey, C. Deaton, T. Cuisset, S. Agewall, K. Dickstein, T. Edvardsen, J. Escaned, B.J. Gersh, P. Svitil, M. Gilard, D. Hasdai, R. Hatala, F. Mahfoud, J. Masip, C. Muneretto, M. Valgimigli, S. Achenbach, J.J. Bax, ESC Scientific Document Group, 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes, Eur. Heart J. 41 (2020) 407–477. https://doi.org/10.1093/eurheartj/ehz425.

[24] R.M. Lang, L.P. Badano, V. Mor-Avi, J. Afilalo, A. Armstrong, L. Ernande, F.A. Flachskampf, E. Foster, S.A. Goldstein, T. Kuznetsova, P. Lancellotti, D. Muraru, M.H. Picard, E.R. Rietzschel, L. Rudski, K.T. Spencer, W. Tsang, J.-U. Voigt, Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging, J. Am. Soc. Echocardiogr. Off. Publ. Am. Soc. Echocardiogr. 28 (2015) 1-39.e14. https://doi.org/10.1016/j.echo.2014.10.003.

[25] P. Elliott, B. Andersson, E. Arbustini, Z. Bilinska, F. Cecchi, P. Charron, O. Dubourg, U. Kühl, B. Maisch, W.J. McKenna, L. Monserrat, S. Pankuweit, C. Rapezzi, P. Seferovic, L. Tavazzi, A. Keren, Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases, Eur. Heart J. 29 (2008) 270–276. https://doi.org/10.1093/eurheartj/ehm342.

[26] E. Fornasa, A. Accardo, M. Cinquetti, M. Merlo, G. Sinagra, HRV spectral and fractal analysis in heart failure patients with different aetiologies, in: Comput. Cardiol. 2014, 2014: pp. 421–424.

[27] Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Circulation. 93 (1996) 1043–1065.
[28] M.A. Woo, W.G. Stevenson, D.K. Moser, R.B. Trelease, R.M. Harper, Patterns of beat-to-beat heart rate variability in advanced heart failure, Am. Heart J. 123 (1992) 704–710. https://doi.org/10.1016/0002-8703(92)90510-3.
[29] T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, Phys. Nonlinear Phenom. (1988). https://doi.org/10.1016/0167-2789(88)90081-4.
[30] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification And Regression Trees, Routledge, Boca Raton, 1984. https://doi.org/10.1201/9781315139470.
[31] W.-Y. Loh, Fifty Years of Classification and Regression Trees, *Int. Stat. Rev.*, vol. 82, no. 3, pp. 329–348, 2014, doi: 10.1111/insr.12016.
[32] A. Urso, A. Fiannaca, M. La Rosa, V. Ravì, R. Rizzo, Data Mining: Prediction Methods, in: S. Ranganathan, M. Gribskov, K. Nakai, C. Schönbach (Eds.), Encycl. Bioinforma. Comput. Biol., Academic Press, Oxford, 2019: pp. 413–430. https://doi.org/10.1016/B978-0-12-809633-8.20462-7.
[33] Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, Elsevier, 2018.
[34] I. Rish, An empirical study of the naive Bayes classifier, in: IJCAI 2001 Workshop Empir. Methods Artif. Intell., 2001: pp. 41–46.
[35] J. Lubsen, J. Pool, E. van der Does, A practical device for the application of a diagnostic or prognostic function, Methods Inf. Med. 17 (1978) 127–129.
[36] M. Možina, J. Demšar, M. Kattan, B. Zupan, Nomograms for Visualization of Naive Bayesian Classifier, in: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Knowl. Discov. Databases PKDD 2004, Springer, Berlin, Heidelberg, 2004: pp. 337–348. https://doi.org/10.1007/978-3-540-30116-5_32.
[37] Nomogram — Orange Documentation v2.7.8, (n.d.). https://docs.biolab.si/orange/2/widgets/rst/classify/nomogram.html (accessed April 1, 2022).
[38] G. Corani, A. Benavoli, A Bayesian approach for comparing cross-validated algorithms on multiple data sets, Mach. Learn. 100 (2015) 285–304. https://doi.org/10.1007/s10994-015-5486-z.
[39] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, B. Zupan, Orange: Data Mining Toolbox in Python, J. Mach. Learn. Res. 14 (2013) 2349–2353.
[40] M.G. Poddar, V. Kumar, Y.P. Sharma, Automated diagnosis of coronary artery diseased patients by heart rate variability analysis using linear and non-linear methods, J. Med. Eng. Technol. 39 (2015) 331–341. https://doi.org/10.3109/03091902.2015.1063721.
[41] H. Li, L. Zhao, Algorithmic Study of the Characteristics of Electrocardiograph Signals in Patients with Coronary Heart Disease, Sci. Program. 2021 (2021) e2304072. https://doi.org/10.1155/2021/2304072.
[42] A.C. Solal, P. Assyag, P. Clerson, C. Contre, M. Guenoun, P. Poncelet, J.F. Thebaut, L. Irina, 092 “Grey Zone” of 40-50% ejection fraction in ambulatory patient with Heart Failure. Who are these patients? Lessons from the DEVENIR study, Arch. Cardiovasc. Dis. Suppl. 2 (2010) 31. https://doi.org/10.1016/S1878-6480(10)70094-X.
[43] M. Pagani, F. Lombardi, S. Guzzetti, G. Sandrone, O. Rimoldi, G. Malfatto, S. Cerutti, A. Malliani, Power spectral density of heart rate variability as an index of sympatho-vagal interaction in normal and hypertensive subjects, J. Hypertens. Suppl. Off. J. Int. Soc. Hypertens. 2 (1984) S383-385.
[44] H.V. Huikuri, T.H. Mäkikallio, Heart rate variability in ischemic heart disease, Auton. Neurosci. Basic Clin. 90 (2001) 95–101. https://doi.org/10.1016/S1566-0702(01)00273-9.