

Article

A Machine Learning Approach for Chronic Heart Failure Diagnosis

Dafni K. Plati ¹, Evanthia E. Tripoliti ¹, Aris Bechlioulis ², Aidonis Rammos ², Iliada Dimou ², Lampros Lakkas ², Chris Watson ^{3,4}, Ken McDonald ⁴, Mark Ledwidge ⁴, Rebabonye Pharithi ⁴, Joe Gallagher ⁴, Lampros K. Michalis ², Yorgos Goletsis ^{1,5}, Katerina K. Naka ² and Dimitrios I. Fotiadis ^{1,*}

- ¹ Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, 45110 Ioannina, Greece; daphni.plati@gmail.com (D.K.P.); etripoliti@gmail.com (E.E.T.); goletsis@uoi.gr (Y.G.)
- ² 2nd Department of Cardiology, Faculty of Medicine, School of Health Sciences, University of Ioannina, 45110 Ioannina, Greece; md02798@yahoo.gr (A.B.); aidrammos@yahoo.gr (A.R.); iliadadimou@gmail.com (I.D.); ftpcavalier52@gmail.com (L.L.); lamprosmichalis@gmail.com (L.K.M.); drkknaka@gmail.com (K.K.N.)
- ³ Wellcome-Wolfson Institute for Experimental Medicine, Queen's University, Belfast BT9 7BL, UK; chris.watson@qub.ac.uk
- ⁴ University College Dublin, National University of Ireland, Belfield, D04 Dublin, Ireland; kenneth.mcdonald@ucd.ie (K.M.); mark.ledwidge@ucd.ie (M.L.); rpharithi@gmail.com (R.P.); jgallagher@ucd.ie (J.G.)
- ⁵ Department of Economics, University of Ioannina, 45110 Ioannina, Greece
- * Correspondence: fotiadis@uoi.gr; Tel.: +30-265-100-9006



Citation: Plati, D.K.; Tripoliti, E.E.; Bechlioulis, A.; Rammos, A.; Dimou, I.; Lakkas, L.; Watson, C.; McDonald, K.; Ledwidge, M.; Pharithi, R.; et al. A Machine Learning Approach for Chronic Heart Failure Diagnosis. *Diagnostics* **2021**, *11*, 1863. <https://doi.org/10.3390/diagnostics11101863>

Academic Editors: Laura Burattini, Cees A. Swenne and Agnese Sbrollini

Received: 25 August 2021
Accepted: 7 October 2021
Published: 10 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The aim of this study was to address chronic heart failure (HF) diagnosis with the application of machine learning (ML) approaches. In the present study, we simulated the procedure that is followed in clinical practice, as the models we built are based on various combinations of feature categories, e.g., clinical features, echocardiogram, and laboratory findings. We also investigated the incremental value of each feature type. The total number of subjects utilized was 422. An ML approach is proposed, comprising of feature selection, handling class imbalance, and classification steps. The results for HF diagnosis were quite satisfactory with a high accuracy (91.23%), sensitivity (93.83%), and specificity (89.62%) when features from all categories were utilized. The results remained quite high, even in cases where single feature types were employed.

Keywords: heart failure; machine learning

1. Introduction

HF is a clinical syndrome of various etiologies, in which the heart cannot pump enough blood to satisfy the metabolic needs of the body [1]. Patients with HF have to deal with changes that severely affect their quality of life. HF is one of the major causes of mortality [2] and the most common cause of hospital admissions in people over 65 years of age [3]. Projections show that the prevalence of HF will increase by 46% from 2012 to 2030, resulting in >8 million people ≥ 18 years of age with HF [4]. Based on the left ventricular ejection fraction (EF), HF patients are classified as HF with reduced EF (HFrEF; EF < 40%), mid-range EF (HFmrEF; EF 40–49%), and preserved EF (HFpEF; EF $\geq 50%$) [5].

The diagnosis of HF can be challenging, especially in the early stages and in patients with HFpEF. Symptoms and signs may be particularly difficult to identify and interpret in obese individuals, in the elderly, and in patients with chronic lung disease [3]. Along with standard laboratory investigations, echocardiogram, electrocardiogram (ECG), and natriuretic peptides are probably the most useful tests for diagnosis in patients with suspected HF. For the management of HF, the European Society of Cardiology (ESC) Guidelines [5] propose a range of approaches, including medical management with renin-angiotensin–aldosterone system therapies, beta blockers, mineralocorticoid receptor an-

tagonists, sodium–glucose co-transporter 2 inhibitors (for HF_rEF) and diuretic therapies, careful management of cardiovascular and other co-morbidities, biomarker monitoring with serial natriuretic peptide measurements, remote monitoring (using an implanted device when indicated), structured telephone support, and multidisciplinary care. For most patients, the standard management of HF involves office-based follow-up 2–12 times a year, while patients are advised to monitor their weight, blood pressure, pulse, diet, and symptoms on a daily basis.

ML offers the potential to improve healthcare efficiency in numerous ways. Prognostic models may empower healthcare experts to select better treatment options for their patients. Additionally, diagnostic models can be used in screening, in risk stratification, and in recommending appropriate tests. This decreases the burden on clinicians, saves resources, and reduces costs. Due to the increased incidence and the large financial costs associated with the management of HF, the diagnosis and treatment of the disease remain extremely important issues.

Several studies have been conducted to build a model that can diagnose HF based on various ML algorithms. Ali et al. [6], Javeed et al. [7], Samuel et al. [8], Mohan et al. [9], and Potter et al. [10] utilized the Cleveland Heart Disease Database that consists of demographics, symptoms, clinical and laboratory values, and electrocardiographic features. Choi et al. [11] detected HF on a multivariate dataset consisting of demographics, habits, clinical and laboratory values, the International Classification of Disease version 9 (ICD-9) codes, information in Current Procedural Terminology (CPT) codes, and medication features. Son et al. [12] tested a rough set (RS)-based model on demographic characteristics and clinical laboratory values. Reddy et al. [13] detected HF_pEF by analyzing medications, demographics, comorbidities, and echocardiographic and ECG features. Masetic et al. [14], Acharya et al. [15], and Ning et al. [16] analyzed ECG signals to detect HF. Lal et al. [17], Wang et al. [18], Chen et al. [19], and Gladence et al. [20] utilized Heart Rate Variability (HRV) measures to diagnose congestive HF. Zheng et al. [21] and Gjoreski et al. [22] suggested a system for chronic HF diagnosis based on the analysis of heart sound characteristics. In Table 1, all studies mentioned in the literature review are presented in detail in order to discriminate between different approaches, methods, and datasets.

Table 1. State of the art in machine learning for HF diagnosis.

Study	Target	Method	Features	Dataset	Measures
Zheng et al. [21] (2015)	Chronic HF diagnosis Healthy vs. chronic HF	Least square-Stacked Support Vector Machine (SVM) model	Cardiac reserve and heart sound characteristics	152 subjects 88 controls 64 chronic with HF	Acc 95.39% Sens 96.59% Spec 93.75%
Masetic et al. [14] (2016)	Congestive HF diagnosis Healthy vs. congestive HF	Decision tree, K-Nearest Neighbors (K-NN), SVM, Neural Network (NN), and Random Forest (RF)	ECG signals	31 subjects 18 with congestive HF 13 controls	RF acc 100%
Choi et al. [11] (2017)	HF diagnosis Healthy vs. HF	Recurrent Neural Network (RNN) models, Logistic Regression (LR), SVM, Multilayer Perceptron (MLP), K-NN	Demographics, habits, clinical and laboratory values, ICD-9 codes, CPT codes, and medications	3884 with HF 28.903 controls	RNN model AUC 77.70%

Table 1. Cont.

Study	Target	Method	Features	Dataset	Measures
Chen et al. [19] (2017)	Congestive HF diagnosis Healthy vs. congestive HF	Deep Neural Network (DNN)	HRV measures based on the RR interval	116 subjects 44 with congestive HF 72 controls	Acc 72.44% Sens 50.39% Spec 84.93%
Samuel et al. [8] (2017)	HF diagnosis Healthy vs. HF	Hybrid decision support method based on artificial neural networks and fuzzy analytic hierarchy process (Fuzzy_AHP) techniques	Demographics, symptoms, clinical and laboratory values, and electrocardiographic results	Cleveland heart disease database 297 subjects 137 with HF 160 controls	Acc 91.10%
Reddy et al. [13] (2018)	HFpEF identification	LR	Medications, demographics, comorbidities, and echocardiographic and ECG features	414 subjects 267 with HFpEF 147 controls	AUC 88.60%
Wang et al. [18] (2019)	Congestive HF diagnosis Healthy vs. congestive HF	Combination of the Long Short-Term Memory (LSTM) network and convolution net architecture	HRV measures based on the RR interval	156 subjects 44 with congestive HF 112 controls	Acc 99.22%
Acharya et al. [15] (2019)	Congestive HF diagnosis Healthy vs. congestive HF	Convolutional neural network (CNN)	ECG signals	73 subjects 15 with congestive HF 58 controls	Acc 98.97% Spec 99.01% Sens 98.87%
Ali et al. [6] (2019)	HF diagnosis Healthy vs. HF	SVM	Demographics, symptoms, clinical and laboratory values, and electrocardiographic results	Cleveland heart disease database 297 subjects 137 with HF 160 controls	Acc 92.22% Sens 100.00% Spec 82.92%
Javeed et al. [7] (2019)	HF diagnosis Healthy vs. HF	Random Search Algorithm (RSA) for feature selection and RF for classification	Demographics, symptoms, clinical and laboratory values, and electrocardiographic results	Cleveland heart disease database 297 subjects 137 with HF 160 controls	Acc 93.33%
Mohan et al. [9] (2019)	HF diagnosis Healthy vs. HF	Hybrid RF	Demographics, symptoms, clinical and laboratory values, and electrocardiographic results	Cleveland heart disease database 297 subjects 137 with HF 160 controls	Acc 88.40% Sens 92.80% Spec 82.60%
Lal et al. [17] (2020)	Congestive HF diagnosis Healthy vs. congestive HF	SVM Gaussian, K-NN, decision tree, SVM RBF, and SVM polynomial	HRV measures	116 subjects 44 with congestive HF 72 controls	SVM Gaussian Acc 88.79% Sens 93.06% Spec 81.82% AUC 95.00%
Gjoreski et al. [22] (2020)	Chronic HF diagnosis Healthy vs. chronic HF	Combination of classic ML and end-to-end Deep Learning (DL)	Heart sound characteristics	947 subjects	Acc 92.90% Sens 82.30% Spec 96.20%

Table 1. Cont.

Study	Target	Method	Features	Dataset	Measures
Potter et al. [10] (2020)	Stage B HF detection	RF	Demographics, symptoms, clinical and laboratory values, and electro- cardiographic results	Cleveland Heart Disease Database 254 subjects as train set (135 with HF, 119 controls) 65 subjects as test set (27 with HF, 38 controls)	AUC 76.00% Sens 93.00% Spec 61.00%
Ning et al. [16] (2020)	Congestive HF diagnosis Healthy vs. congestive HF	Hybrid DL algorithm that is composed of a CNN and a recursive NN	ECG signals	33 subjects 15 chronic HF subjects 18 controls	Acc 99.93% Sens 99.85% Spec 100%

All previous works focus on classification between HF and non-HF, using various methods, datasets, and features. Such a classification, although very useful for an automated diagnosis system, provides limited support to an experienced clinician that lacks the ability to perform laboratory tests and echocardiogram due to various logistic reasons [23,24]. In the present study, we propose a methodology to diagnose HF; its main characteristic is that the models are based on various combinations of features, using the clinical approach followed by clinicians, based on current guidelines [5]. In order to examine how each feature type contributes to the diagnosis, initially, our models were built by utilizing only clinical features, i.e., features that can be collected by all clinicians without performing laboratory tests or echocardiogram, such as the patient's medical history, results from the physical examination, symptoms, comorbidities, demographic information, etc. Subsequently, various combinations of features are assessed by adding additional types of features: Clinical features and natriuretic peptides, clinical and echocardiographic features, echocardiographic features exclusively, and finally all features combined. In this way, we examined how—through the application of machine learning—the different types of features proposed in the guidelines can be applied for HF diagnosis. The results can also be of special value in cases where some of the features cannot be easily obtained.

2. Materials and Methods

2.1. The Dataset

Data were provided by the University College Dublin (UCD), Ireland (410 subjects), and the 2nd Department of Cardiology of the University Hospital of Ioannina (77 subjects). The total number of subjects was 487 (260 without HF, 180 with chronic HF, and 47 with acute HF). As is common in ML studies, the maximum available data were used (available patient data from patients having accepted informed consent). In order to train and test ML algorithms, both kinds of subjects are needed, i.e., HF and controls. This allows algorithms to be trained for both cases and to be able to correctly classify any new case. The same applies for testing; testing needs to be done for both cases (HF/non-HF). In our case, we had 227 HF patients and 260 controls. Patients were diagnosed with HF by clinical experts. This diagnosis was based on the patients' physical examinations, along with standard laboratory investigations, echocardiograms, ECGs, and the measurement of natriuretic peptides. The features recorded for each patient were grouped into the following categories: General demographic data, classical cardiovascular risk factors, personal history of cardiovascular disease, other diseases, lifestyle/habits, medications, symptoms, physical examination, laboratory findings, and echocardiographic features (Table 2).

Table 2. Initial dataset provided.

Category	Description
General demographic data	Age and gender
Classical cardiovascular risk factors	Hypertension and diabetes mellitus
Personal history of cardiovascular disease	Device, myocardial infarction (MI), coronary artery disease (CAD), angina, peripheral vascular disease, any arrhythmia (Arr), paroxysmal atrial fibrillation (Afib), and stroke
Other diseases	Arthritis, chronic obstructive pulmonary disease, cancer, asthma, gout
Lifestyle/habits	smoking, and physical activity
Medications	Mineralocorticoid receptor antagonists (MRAs), diuretics (loop or thiazide diuretic), calcium channel blocker (CCB), statin, antiplatelet, renin angiotensin aldosterone system (RAAS), beta blocker (BB), oral anticoagulant (OAC), other lipid-lowering drugs (LipD), alpha blocker, digoxin, insulin, warfarin, nitrate, diabetes drugs, and ivabradine
Symptoms	Dyspnea, orthopnea, NYHA classes I–IV, and paroxysmal nocturnal dyspnea
Physical examination	Weight, height, body mass index (BMI), murmurs, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), pulse, crackles, oedemas, JVP distension, and body surface area
Laboratory findings	BNP, Na, K, Ca, Cl, urea, creatinine, eGFR, full blood count including WBC, full blood count including Hb, platelet count, total cholesterol, HDL, LDL, triglycerides, and glucose (non-fasting)
Echocardiographic parameters	Interventricular septal thickness at end-diastole (IVS), posterior wall thickness at end diastole (PW), left ventricular internal dimension in diastole (LVIDd), LV mass, left ventricular mass index (LVMI), left atrial volume (average 4ch and 2ch) (LAVI), left atrial (LA) dimension (mm), peak E-value, peak A-value, early filling (E wave)/late diastolic filling (A wave) ratio (E/A), mitral annular velocity (E'), early filling (E wave)/mitral annular velocity (E/E'), E deceleration time, ejection fraction (EF), diastolic biventricular inner dimension, estimation of any valvular disease, right ventricular systolic pressure, and pulmonary artery systolic pressure; classification of HF phenotype into: HF _r EF, HF _{mr} EF, and HF _p EF

Certain features were removed due to overlapping information, i.e., interventricular septal thickness and posterior wall thickness at the end diastole had similar information to the left ventricular mass index values. NYHA classification and dyspnea symptoms were recorded in all patients, but were not included in the HF classification analysis. Rhythm device information was also excluded from the final analysis, since the target population who may benefit from the application of these models will not have implanted defibrillators or resynchronization therapy (these therapies are applied to already diagnosed symptomatic

NYHA II–IV HF patients). Finally, the classification of HF phenotype (HFpEF, HFrEF, or HFmrEF) did not fit in our analysis, since the models were to be used for HF diagnosis in previously undiagnosed patients. Medications were not included in our final dataset, since they can be considered either as a determinant or as an indicator of a patient’s condition. The resulting dataset is depicted in Table 3.

Table 3. Dataset after deleting features.

Category	Description
General demographic data	Age and gender
Classical cardiovascular risk factors	Hypertension and diabetes mellitus
Personal history of cardiovascular disease	MI, CAD, angina, peripheral vascular disease, Arr, Afib, and stroke
Other diseases	Arthritis, chronic obstructive pulmonary disease, cancer, asthma, and gout
Lifestyle/habits	Smoking and physical activity
Symptoms	Orthopnea, Paroxysmal Nocturnal Dyspnea
Physical examination	Weight, height, BMI, murmurs, SBP, DBP, HR, pulse, crackles, edemas, JVP distension, and body surface area
Laboratory findings	BNP, Na, K, Ca, Cl, urea, creatinine, eGFR, full blood count including WBC, full blood count including Hb, platelet count, total cholesterol, HDL, LDL, triglycerides, and glucose (non-fasting)
Echocardiographic parameters	LVIDd, LV mass, LVMI, LAVI, left atrial dimension (mm), peak E-value, peak A-value, EA, mitral annular velocity, Ee, E deceleration time, EF, diastolic biventricular inner dimension, estimation of any valvular disease, right ventricular systolic pressure, and pulmonary artery systolic pressure

2.2. The Proposed Methodology

The proposed methodology consists of three basic stages: Preprocessing, feature selection, and classification. The preprocessing pipeline includes the removal of features with $\geq 50\%$ missing values. Discrete features with unbalanced distribution of values are also removed and outliers and typos (e.g., 4,5 is recorded instead of 4.5) per feature are detected and corrected. Furthermore, the class imbalance problem is handled by applying the undersampling method to the dataset, which is quite common for dealing with this issue [25]. This method produces a random subsample with a given spread between class frequencies. The maximum “spread” between the rarest and most common class is specified. In this study, a random undersample of the majority class is followed, so that all classes have the same number of instances. The 19 features retained after removing features with $\geq 50\%$ missing values and discrete features with unbalanced distribution of values and utilized for the diagnosis of HF are presented in Table 4. For better understanding of the individual contribution of each feature, we calculated the information gain [26] metric as presented in Table A2 of Appendix B.

Table 4. Features for HF diagnosis.

Category	Description
General demographic data	Age and gender
Classical cardiovascular risk factors	Hypertension
Personal history of cardiovascular disease	MI, CAD, and any arrhythmia (Arr) or paroxysmal atrial fibrillation (Afib) combined as Arr-Afib
Physical examination	BMI, SBP, DBP, and HR
Laboratory findings	BNP
Echocardiographic parameters	LVIDd, LVMI, LAVI, EA, E deceleration time, Ee, EF, and peak E-value

At the second stage, feature selection is applied to all features from all categories and the subset of features retained is used for the classification process. In this study, feature selection methods are employed to assess the predictive ability of feature subsets and the degree of redundancy among them, preferring sets of features that are highly correlated with the class but with low intercorrelation [27].

Finally, at the classification stage, different classifiers (i.e., decision tree, RF, rotation forest (ROT), naive Bayes (NB), K-NN, SVM, logistic model tree (LMT), and Bayes network (BN)) are applied to the reduced feature subset. Then, 10-fold cross-validation is applied for the evaluation of the classifiers, which is a statistical method for evaluating and comparing learning algorithms by dividing data into two segments: One used to learn or train a model and the other used to validate the model [28]. The proposed methodology is presented in Figure 1. The results are expressed in terms of accuracy, sensitivity, and specificity.

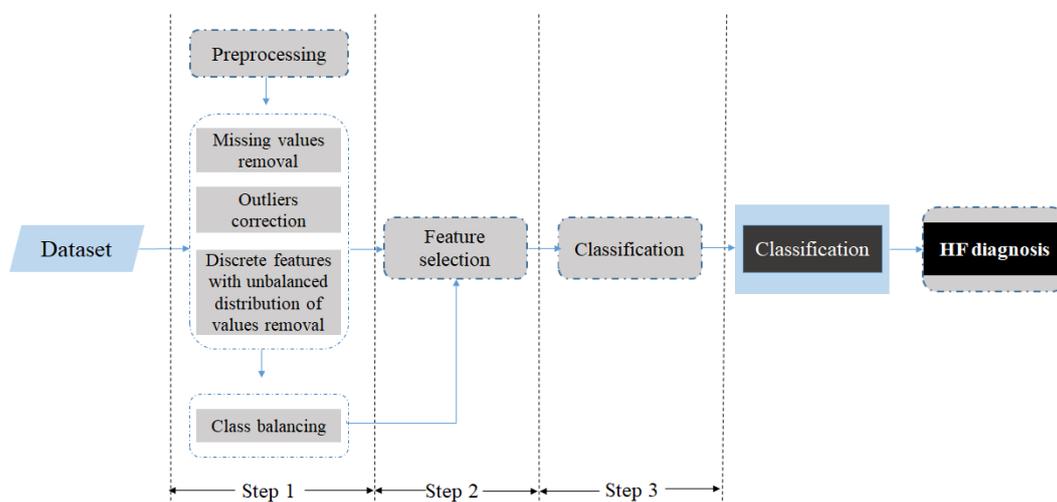


Figure 1. Methodology.

The aforementioned methodology was applied for the following combination of features (Figure 2).

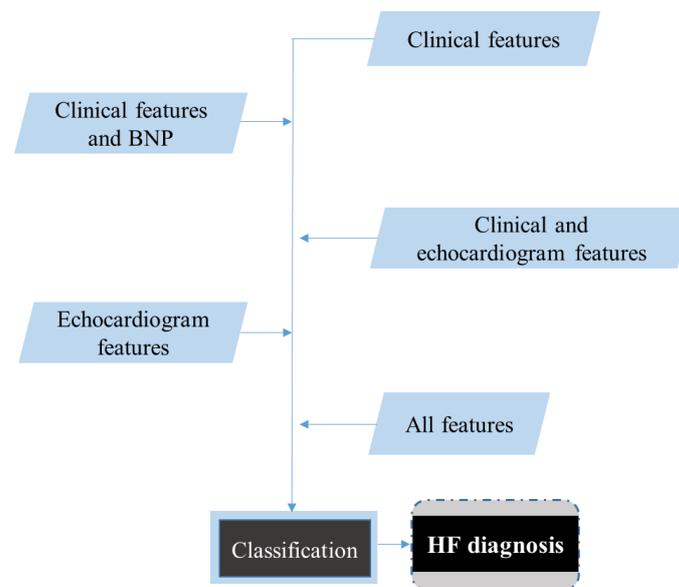


Figure 2. Classification for various sets of features.

3. Results

The mean age of the subjects in the dataset was 69 years and the median age was 71 years. Regarding HF patients, the mean age was 72 and the median was 74 years, while for subjects without HF, the mean age was 67 and median was 68 years. The dataset consisted of 260 male and 224 female subjects. The dataset with documented HF consisted of 73 females and 154 males, while the non-HF dataset consisted of 151 female and 106 male subjects (for three subjects, gender information was unavailable). Furthermore, regarding the subjects with HF, 68 were classified as HFpEF, while 60 as HFrEF and 88 as HFmrEF (for 11 subjects, the classification according to ejection fraction was missing, since the ejection fraction was unavailable).

For the HF diagnosis, all subjects with acute HF and subjects with NYHA classification III–IV were removed, as diagnosis of HF in patients with severe symptoms is not challenging. Thus, the dataset for HF diagnosis consisted of 422 subjects (260 without HF and 162 with chronic HF).

The results for the HF diagnosis problem with feature selection are depicted in Table 5.

Table 5. HF diagnosis classification results.

Features Type	Classifier	Accuracy %	Sensitivity %	Specificity %
Clinical features	LMT	84.12	82.10	85.38
Clinical features and BNP	LMT	88.15	85.80	89.62
Clinical and echocardiographic features	ROT	90.76	93.21	89.23
Echocardiographic features	ROT	87.91	90.74	86.15
All features	ROT	91.23	93.83	89.62

Our approach achieved the highest results in terms of accuracy using mostly the LMT and ROT classifiers with different combinations of features included. The accuracy values ranged from 84.12% in models using only clinical features to 91.23% in models using all features combined.

The optimal features that were retained after the feature selection procedure in various models with different feature combinations are presented in Table 6. All retained features showed statistically significant ($p < 0.01$) correlations (Figure A1 and Tables A3 and A4) with the diagnosis of HF.

Table 6. Retained features for HF diagnosis, all possible features set.

Feature Set	Retained Features
Clinical features	Hypertension, Arr-Afib, CAD, and SBP
Clinical features and BNP	Hypertension, Arr-Afib, CAD, SBP, and BNP
Clinical and echocardiogram features	Hypertension, Arr-Afib, CAD, SBP, EF, LAVI, LVMI, E/E', and E deceleration time
Echocardiogram features	EF, LAVI, LVMI, and E deceleration time
All features	Hypertension, Arr-Afib, LAVI, LVMI, CAD, BNP, SBP, and EF

4. Discussion

Data-driven approaches for the optimization of population health management are continuously growing and may prove valuable in modern healthcare models, especially for highly prevalent and costly diseases such as HF. The present study made another considerable contribution toward the development of such an approach for the diagnosis of HF in symptomatic patients with risk factors based on simple clinical data, as well as natriuretic peptides and echocardiographic indices (suggested by ESC guidelines) with the use of various machine learning techniques. The results for HF diagnosis were quite high in terms of accuracy (91.23%), as well as in terms of sensitivity (93.83%) and specificity (89.62%), confirming the classification power of ML approaches. Furthermore, our model achieved high accuracy, even when only the clinical features were used for classification (84.12%), which can prove to be of great value for an initial screening in settings where laboratory tests are not available. We also noticed that the addition of BNP to clinical features increased the accuracy of HF diagnosis (88.15%), as expected based on the well-established value of natriuretic peptides in the diagnosis of HF [5]. Furthermore, the combination of clinical and echocardiographic features for the classification of HF diagnosis also resulted in increased accuracy compared to clinical features alone (accuracy 90.76%). This finding re-emphasizes the diagnostic value of echocardiography that, even without natriuretic peptides, can establish the diagnosis of HF in the majority of patients with suspected non-acute HF [5]. A small difference in terms of accuracy between models using BNP or echocardiographic parameters was observed; whether these differences are of clinical importance is not known.

Furthermore, in every classification model, our method finally utilized a smaller feature subset. This may indicate that a small number of clinical, biochemical, and echocardiographic parameters are needed for reaching a diagnosis of HF, which corresponds to less time and cost. From a clinical point of view, the current study suggests that the identification of a few classic risk factors (e.g., hypertension) and common cardiovascular diseases (e.g., coronary artery disease and atrial fibrillation) are most important in the diagnosis of HF. Regarding the laboratory tests taken into account, BNP (and its N-terminal counterpart NTproBNP) is the biomarker most widely used. It is secreted mainly by the ventricular heart muscle and causes natriuresis, diuresis, and smooth muscle relaxation. It is increased in HF of all etiologies [29], both in acute and chronic settings. We might have also used troponin T (TnT), a polypeptide forming part of the contractile apparatus of the striated muscle, which is the best laboratory parameter in the early diagnosis of acute myocardial infarction and also has a predictive role in various diseases of the cardiovascular system, such as HF and/or in hemodynamic instability [30,31]. The latest guidelines by the ESC recommend the use of troponin for the exclusion of an acute coronary syndrome and suggest it for estimating the risk of myocardial damage in other situations such as

hereditary muscle diseases (dilated, hypertrophic, and arrhythmogenic cardiomyopathy), cardio-toxic medications used in oncology, myocarditis, and atrial diseases [5]. However, in chronic HF, TnT tends to be less marked and it is used as incremental information to natriuretic peptides [32]. For simplicity reasons we included only the biomarker that was incorporated in the diagnostic algorithm for HF proposed by the ESC [5].

Besides natriuretic peptides, the use of only a few structural echocardiographic indices (such as LV mass, left atrial size, and LV ejection fraction) may be also valuable for HF diagnosis. In this way, ML may assist in the isolation of high-risk features and thus provide an appropriate phenotyping that may improve the detection accuracy of HF. These methods may also lead to greater insights into the pathophysiological pathways underlying the development of HF and the design of future clinical studies that will validate the clinical importance of our findings.

Relevant approaches in the literature provide a method for detecting HF based on several feature types with an accuracy ranging from 72.44% to 99.22%. Our method cannot be directly compared with those utilizing ECG signals [14–16], HRV measures [17–19], or heart sound characteristics [21,22], but with those that utilize the Cleveland Heart Disease Database [6–10], which is a dataset that resembles ours, and studies that use a multivariate dataset [11–13]. It should be noted that the present study utilized a larger dataset (422 instances) compared to the studies that utilized the Cleveland Heart Disease Database. Moreover, in our approach, the medications were not finally considered in the feature set. If we included medications, the obtained results would further increase (the ROT classifier achieved 93.36% accuracy, 95.70% sensitivity, and 91.90% specificity). Still, as the addition of medications might introduce a kind of bias, this approach was not selected. We also tested whether CAD and Arr-Afib could be omitted, as they are not necessarily known or easy to determine during a consultation. It seems that these two features can slightly contribute to the performance of our classifier. In more detail, by omitting these two features the evaluation metrics (namely, the accuracy, sensitivity, and specificity) changed from 91.23%, 93.83%, and 89.62% to 90.28%, 94.00%, and 85.00% (Table A1). On the other side, this is an indication that our approach works adequately, even without these hard-to-obtain features.

Moreover, we excluded several features (NYHA class, device, dyspnea, and HF phenotype) and subjects with acute HF and NYHA classes III–IV as they could be indicative of HF presence. In our study, feature selection was applied, concluding to a smaller feature set where all retained features were significantly correlated with the class (Tables A2 and A3); even with a smaller feature set the achieved results were high. This study provides an automated diagnostic tool with high accuracy for detecting the presence of HF, even in cases when limited tests (echocardiogram and laboratory tests) are offered. Additionally, it can be valuable in cases when multiple co-morbidities occur and can offer the clinical expert a further aid in the diagnosis of HF.

Limitations: Although the current study was performed with one of the largest datasets compared to the literature, the incorporation of the proposed approach in a Clinical Decision Support System used in actual clinical practice requires extensive testing and validation with a larger and more diverse dataset.

5. Conclusions

In the present study, we developed a method approach able to diagnose the presence of HF based on ML techniques. This study is quite innovative, because we simulated the clinical procedure and investigated the impact of different feature types on the classification accuracy. The results for the HF diagnosis, when all available feature types were utilized for classification, were high in terms of accuracy (91.23%), sensitivity (93.83%), and specificity (89.62%). Efficiency is supported as a limited feature set is selected through feature selection, minimizing the need for diagnostic tests. Moreover, even without the whole feature set, our approach provides quite high results; the results remain high even when only clinical features are used. This provides opportunity to clinicians that do not have the opportunity

to perform laboratory tests or echocardiograms to diagnose HF quite accurately without necessarily needing the input of additional tests.

Author Contributions: Conceptualization, D.I.F., Y.G., K.K.N. and L.K.M.; methodology, D.K.P., E.E.T., Y.G. and K.K.N.; validation K.K.N., A.B., A.R., I.D., L.L. and L.K.M.; resources, A.B., A.R., I.D., L.L., C.W., K.M., M.L., R.P. and J.G.; data curation, D.K.P. and E.E.T.; writing—original draft preparation, D.K.P.; writing—review and editing, D.K.P., E.E.T., Y.G., K.K.N. and C.W.; visualisation, D.K.P.; supervision, D.I.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the KardiaTool project (<http://www.kardiatooleu/>) that received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 768686. This article reflects only the authors’ views. The Commission is not responsible for any use that may be made of the information it contains.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics and Medical Research Committee of St Vincent’s University Hospital in Dublin (protocol code HBT-STOPHF-PTCL-01 (12/08/2004), HBT-BEMHF-PTCL-01 (11/06/2012)) and by The board of the University Hospital of Ioannina (16/9-5-2019 (9-5-2019)).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

HF Diagnosis Classification Results without CAD and Arr-Afib Features

As CAD and Arr-Afib are not necessarily known or easy to determine during a consultation, we also tested our models without these two features. The results are presented in Table A1.

Table A1. HF diagnosis classification results without CAD and Arr-Afib features.

Features Type	Classifier	Accuracy %	Sensitivity %	Specificity %
Clinical features	NB	75.36	80.70	67.30
Clinical features and BNP	ROT	86.02	92.80	77.50
Clinical and echocardiographic features	ROT	88.39	92.70	82.30
All features	ROT	90.28	94.00	85.00

Appendix B

Contribution of Each Individual Feature to the Predicted Outcome

The information gain calculates the reduction in entropy from transforming a dataset in some way and is used for determining the best features that render the maximum information about a class. In Table A2, the features are ranked based on their information gain, which is also given.

Table A2. Features' information gain.

Ranking	Feature	Information Gain
1	BNP	0.4202
2	Arr_Afib	0.3076
3	LAVI	0.302
4	EF	0.2813
5	DBP	0.131
6	LVMI	0.1181
7	CAD	0.1143
8	SBP	0.1129
9	Hypertension	0.0905
10	Peak_E_Value	0.066
11	Age	0.0544
12	Gender	0.0489
13	Ee	0.0405
14	MI	0.0393
15	LVIDd	0.0312
16	E Deceleration Time	0.02
17	HR	0
18	BMI	0
19	Ee	0

In Figure A1, the correlations among all features, as well as the correlation of each feature with the class (HF outcome), are presented.

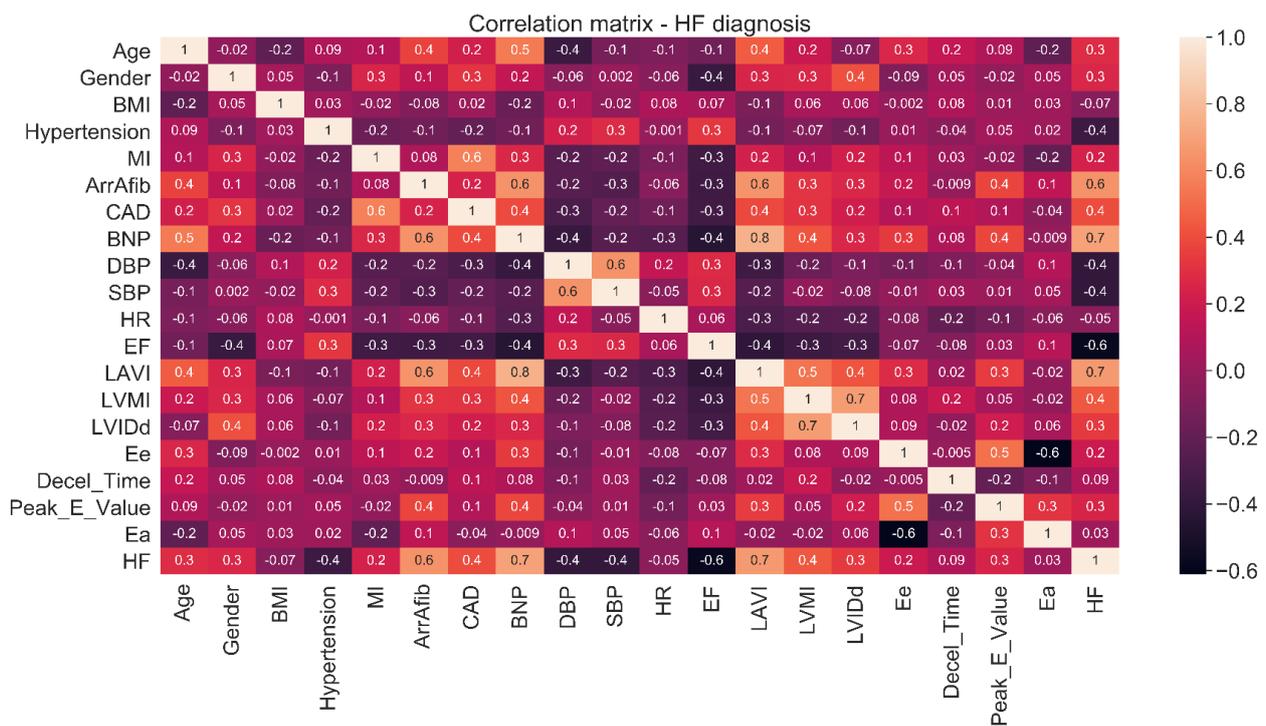


Figure A1. Correlation matrix.

In Table A3, the correlation of each feature with the class (HF outcome) is presented.

Table A3. Features' correlation with class.

Feature	Correlation with Class
Age	0.3
Gender	0.3
BMI	-0.07
Hypertension	-0.4
MI	0.2
Arr-Afib	0.6
CAD	0.4
BNP	0.7
SBP	-0.4
DBP	-0.4
HR	-0.05
EF	-0.6
LAVI	0.7
LVMi	0.4
LVIDd	0.3
Ee	0.2
E Deceleration time	0.09
Peak_E_Value	0.3
EA	0.03

For the retained features, the chi-square test results and the p -values are presented in Table A4.

Table A4. Chi-square and p -value for the retained features.

	Chi-Square Test	p -Value
LAVI	12.791	<0.001
LVMi	8.467	<0.001
SBP	-7.987	<0.001
BNP	14.2434	<0.001
EF	-11.659	<0.001
HyperT	53.74	<0.001
Arr-Afib	172.51	<0.001
CAD	67.11	<0.001

Appendix C

Results from the Logistic Regression for Clinical Features

The results after the application of logistic regression for the clinical features are presented in Table A5, where the results from the LMT classifier (that were the best results) are also depicted.

Table A5. Results from the logistic regression for clinical features.

	Accuracy %	Sensitivity %	Specificity %
Logistic regression	84.12	88.10	78.10
LMT	84.12	82.10	85.38

References

- Braunwald, E.; Zipes, D.P.; Libby, P.; Bonow, R.O. *Heart Disease: A Textbook of Cardiovascular Medicine*, 7th ed.; Elsevier Science: Philadelphia, PA, USA, 2004.
- Sabate, E. *Adherence to Long-Term Therapies, Evidence for Action*; World Health Organization: Geneva, Switzerland, 2003.
- Franzén, K.; Saveman, B.-I.; Blomqvist, K. Predictors for Health Related Quality of Life in Persons 65 Years or Older with Chronic Heart Failure. *Eur. J. Cardiovasc. Nurs.* **2007**, *6*, 112–120. [[CrossRef](#)]
- Mozaffarian, D.; Benjamin, E.J.; Go, A.S.; Arnett, D.K.; Blaha, M.J.; Cushman, M.; de Ferranti, S.; Després, J.-P.; Fullerton, H.J.; Howard, V.J.; et al. Heart Disease and Stroke Statistics—2015 Update. *Circulation* **2015**, *131*, e29–e322. [[CrossRef](#)]
- McDonagh, T.A.; Metra, M.; Adamo, M.; Gardner, R.S.; Baumbach, A.; Böhm, M.; Burri, H.; Butler, J.; Čelutkienė, J.; Chioncel, O.; et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur. Heart J.* **2021**, *42*, 3599–3726. [[CrossRef](#)] [[PubMed](#)]
- Ali, L.; Niamat, A.; Khan, J.A.; Golilarz, N.A.; Xingzhong, X.; Noor, A.; Nour, R.; Bukhari, S.A.C. An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure. *IEEE Access* **2019**, *7*, 54007–54014. [[CrossRef](#)]
- Javeed, A.; Zhou, S.; Yongjian, L.; Qasim, I.; Noor, A.; Nour, R. An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection. *IEEE Access* **2019**, *7*, 180235–180243. [[CrossRef](#)]
- Samuel, O.W.; Asogbon, G.M.; Sangaiah, A.K.; Fang, P.; Li, G. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Syst. Appl.* **2017**, *68*, 163–172. [[CrossRef](#)]
- Mohan, S.; Thirumalai, C.; Srivastava, G. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. *IEEE Access* **2019**, *7*, 81542–81554. [[CrossRef](#)]
- Potter, E.L.; Rodrigues, C.; Ascher, D.; Marwick, T.H. Machine Learning Applied to Energy Waveform Ecg for Prediction of Stage B Heart Failure in the Community. *J. Am. Coll. Cardiol.* **2020**, *75*, 1894. [[CrossRef](#)]
- Choi, E.; Schuetz, A.; Stewart, W.F.; Sun, J. Using recurrent neural network models for early detection of heart failure onset. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 361–370. [[CrossRef](#)]
- Son, C.-S.; Kim, Y.-N.; Kim, H.-S.; Park, H.-S.; Kim, M.-S. Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. *J. Biomed. Inform.* **2012**, *45*, 999–1008. [[CrossRef](#)]
- Reddy, Y.N.V.; Carter, R.E.; Obokata, M.; Redfield, M.M.; Borlaug, B.A. A Simple, Evidence-Based Approach to Help Guide Diagnosis of Heart Failure With Preserved Ejection Fraction. *Circulation* **2018**, *138*, 861–870. [[CrossRef](#)]
- Masetic, Z.; Subasi, A. Congestive heart failure detection using random forest classifier. *Comput. Methods Programs Biomed.* **2016**, *130*, 54–64. [[CrossRef](#)]
- Acharya, U.R.; Fujita, H.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M.; Tan, R.S. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Appl. Intell.* **2019**, *49*, 16–27. [[CrossRef](#)]
- Ning, W.; Li, S.; Wei, D.; Guo, L.Z.; Chen, H. Automatic Detection of Congestive Heart Failure Based on a Hybrid Deep Learning Algorithm in the Internet of Medical Things. *IEEE Internet Things J.* **2021**, *8*, 12550–12558. [[CrossRef](#)]
- Lal, H.; Wajid, A.; Ishtiaq, K.; Monagi, A.; Jalal, A. Detecting Congestive Heart Failure by Extracting Multimodal Features and Employing Machine Learning Techniques. *BioMed Res. Int.* **2020**, *6*, 1–19.
- Wang, L.; Zhou, X. Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals. *Sensors* **2019**, *19*, 1502. [[CrossRef](#)]
- Chen, W.; Liu, G.; Su, S.; Jiang, Q.; Nguyen, H. A CHF detection method based on deep learning with RR intervals. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 3369–3372.
- Gladence, L.M.; Ravi, T.; Karthi, M. An enhanced method for detecting congestive heart failure—Automatic Classifier. In Proceedings of the 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 8–10 May 2014; pp. 586–590.
- Zheng, Y.; Guo, X.; Qin, J.; Xiao, S. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Comput. Methods Programs Biomed.* **2015**, *122*, 372–383. [[CrossRef](#)]
- Gjoreski, M.; Gradisek, A.; Budna, B.; Gams, M.; Poglajen, G. Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure From Heart Sounds. *IEEE Access* **2020**, *8*, 20313–20324. [[CrossRef](#)]
- From, A.M.; Lam, C.S.P.; Pitta, S.R.; Kumar, P.V.; Balbissi, K.A.; Booker, J.D.; Singh, I.M.; Sorajja, P.; Reeder, G.S.; Borlaug, B.A. Bedside Assessment of Cardiac Hemodynamics: The Impact of Noninvasive Testing and Examiner Experience. *Am. J. Med.* **2011**, *124*, 1051–1057. [[CrossRef](#)] [[PubMed](#)]

24. Thibodeau, J.; Drazner, M.H. The Role of the Clinical Examination in Patients With Heart Failure. *JACC Heart Fail.* **2018**, *6*, 543–551. [[CrossRef](#)] [[PubMed](#)]
25. Luengo, J.; Fernández, A.; García, S.; Herrera, F. Addressing data complexity for imbalanced data sets: Analysis of SMOTE-based oversampling and evolutionary undersampling. *Soft Comput.* **2010**, *15*, 1909–1936. [[CrossRef](#)]
26. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
27. Hall, M. *Correlation-Based Feature Selection for Machine Learning*; University of Waikato: Hamilton, New Zealand, 1998.
28. Marcot, B.G.; Hanea, A.M. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput. Stat.* **2021**, *36*, 2009–2031. [[CrossRef](#)]
29. Zaphiriou, A.; Robb, S.; Murray-Thomas, T.; Mendez, G.; Fox, K.; McDonagh, T.; Hardman, S.M.; Dargie, H.J.; Cowie, M.R. The diagnostic accuracy of plasma BNP and NTproBNP in patients referred from primary care with suspected heart failure: Results of the UK natriuretic peptide study. *Eur. J. Heart Fail.* **2005**, *7*, 537–541. [[CrossRef](#)] [[PubMed](#)]
30. Duchnowski, P.; Hryniewiecki, T.; Małgorzata, K.; Kuśmierczyk, M.; Szymański, P. High-sensitivity troponin T is a prognostic marker of hemodynamic instability in patients undergoing valve surgery. *Biomark. Med.* **2018**, *12*, 1303–1309. [[CrossRef](#)] [[PubMed](#)]
31. Duchnowski, P.; Hryniewiecki, T.; Kuśmierczyk, M.; Szymański, P. High-Sensitivity Troponin T Predicts Postoperative Cardiogenic Shock Requiring Mechanical Circulatory Support in Patients With Valve Disease. *Shock* **2020**, *53*, 175–178. [[CrossRef](#)] [[PubMed](#)]
32. Eggers, K.M.; Lindahl, B. Application of Cardiac Troponin in Cardiovascular Diseases Other Than Acute Coronary Syndrome. *Clin. Chem.* **2017**, *63*, 223–235. [[CrossRef](#)] [[PubMed](#)]