

RESEARCH ARTICLE

A New Medical Decision Support System for Diagnosing HFrEF and HFpEF Using ECG and Machine Learning Techniques

PINAR ÖZEN KAVAS¹, (Member, IEEE), MEHMET RECEP BOZKURT², İBRAHİM KOCAYİĞİT³, AND CAHİT BİLGİN³

¹Faculty of Engineering, Computer Engineering, Kütahya Dumlupınar University, Kütahya 43100, Turkey

²Faculty of Engineering, Electrical-Electronics Engineering, Sakarya University, Sakarya 54050, Turkey

³Faculty of Medicine, Sakarya University, Sakarya 54050, Turkey

Corresponding author: Pinar Özen Kavvas (pinar.ozen@dpu.edu.tr)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Sakarya University Faculty of Medicine Dean's Office under Application No. 16214662/050.01.04/123.

ABSTRACT As heart failure (HF) appears to be a growing epidemic, no case should be overlooked in the diagnosis of HF. Two subtypes of HF by left ventricular ejection fraction (LVEF) are HF with reduced ejection fraction (HFrEF) ($LVEF \leq 40\%$) and HF with preserved ejection fraction (HFpEF) ($LVEF \geq 50\%$). HFrEF is easier to diagnose. However, the diagnosis of HFpEF is more complex and difficult even for specialists. The diagnosis of HFpEF is a problem that is being tried to be solved in medicine. Since LVEF appears normal ($LVEF \geq 50\%$ also in healthy individuals), HFpEF can be confused with chest diseases due to some similar symptoms. The diagnosis of HF subtypes is ideally made using echocardiography. Echocardiography should be performed in all patients with HF; however, it is expensive and requires specialists. Even in high-resource regions, this test is not always performed, and treatment may need to be initiated before the echocardiographic data are obtained. For such situations, economical and practical systems are required. In this study, a medical decision support system was developed to detect HFrEF and HFpEF cases using only 3-lead ECG. From the ECG data of 61 volunteers, 37 features were extracted, of which 16 were Yule-Walker and Burg's method parameters, and 21 were in the time domain. Consequently, 37 features were reduced by feature selection and triple classification was performed with only 4 features with maximum accuracy. This study aimed to determine whether the individuals with HF symptoms were HFrEF, HFpEF, or healthy. Four machine learning algorithms were used for classification. The best classification accuracy rate was 100% for k-NN, and significant results were also obtained from the other three algorithms: SVMs, Decision Trees, and Ensemble Bagged Trees.

INDEX TERMS Artificial intelligence, classification, electrocardiography, heart failure, HFpEF, HFrEF, machine learning.

I. INTRODUCTION

Ejection fraction (EF) is the amount of blood ejected from a ventricle with each beat of the heart. It can be formulated as $(\text{Pulse Volume}) / (\text{End-Diastolic Volume})$. Left Ventricular Ejection Fraction (LVEF) is a measure of pumping efficiency into the systemic circulation, while right ventricular

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino¹.

ejection fraction is a measure of pumping efficiency into the pulmonary circulation. EF is usually measured by echocardiography and is used as an overall measure of a person's heart function. LVEF is usually low in patients with systolic heart failure, and it is an important determinant of the severity of systolic heart failure. Unlike the heart rate in a healthy person, which can be high or low and vary in its daily course, low LVEF is always associated with disease [1].

Heart failure (HF) can be defined as a functional or structural disorder that induces the heart to be unable to supply sufficient oxygen to the tissues [2]. Clinical HF is described in the European HF Guidelines as a clinical syndrome arising from a functional or structural disorder in the heart, and patients have typical signs and symptoms (signs such as ankle swelling, shortness of breath, and fatigue and symptoms such as apex beat displacement, pulmonary crepitation, and increased jugular vein pressure).

As heart systolic dysfunction increases, LVEF gradually decreases, and end-systolic and end-diastolic volumes generally increase. LVEF is important not only as a manifestation of the disease but also because in most clinical studies patients are identified by LVEF.

Basic clinical studies with patients with systolic HF or reduced EF (HFrEF) generally included patients with LVEF < 40% and currently proven treatments are effective only in this group of patients. On the other hand, studies have also been conducted with HF patients with LVEF > 40-45% and no other cardiac disorders. In some of these patients, as well as LVEF is completely normal (usually > 50%), there is no significant reduction in systolic function either. So, the term HF with preserved ejection fraction (HFpEF) was improved to define these patients [3], [4].

The underlying pathophysiological failure in HFpEF patients is thought to be LV diastolic dysfunction. No single echocardiographic parameter is sufficiently precise and reproducible to diagnose LV diastolic dysfunction. So, an exhaustive echocardiographic examination with exactly correlated two-dimensional and Doppler data are suggested [3], [5], [6].

Doppler Ultrasound devices are expensive, and the echocardiographic examination requires an expert and is laborious. HFpEF is a complicated syndrome that may ensue from functional and structural cardiac disorders rather than only one disease presence, correct diagnosis can be difficult even for HF specialists [7].

In the literature, there are studies using artificial intelligence (AI) and machine learning (ML) methods for various heart failure classifications [8], [9], [10], [11], [12], [12], [13], [14], [15]. These studies generally focused on HFrEF, some studied only HFpEF. Classifications were made for two classes. The features used for classification were demographic information such as age, height, and weight, or more laborious features such as those obtained from echocardiography, magnetic resonance imaging, or invasive methods. In recent years, AI algorithms were developed to detect heart failure using ECG. In these studies, algorithms with AI methods were developed for conditions such as aortic stenosis, anemia, atrial fibrillation, cardiac contractile dysfunction, congestive heart failure, and arrhythmia prediction using ECG [12], [16], [17], [18], [19], [20], [21], [22]. In these studies, 12-lead ECG data were used generally.

By applying AI to ECG, it was shown that subtle changes in QRS could be correlated with heart functions such as myocardial fibrosis, congestive HF, the efficacy of diuresis

treatment, etc., and faster and less costly evaluation could be made [23]. Studies were carried out to predict HF-related death and hospital readmission using AI and ML algorithms [24]. In some of those studies, echocardiographic data [25], or many parameters such as clinical phenotyping, laboratory, ECG, and echocardiography were used simultaneously [26]. Congestive heart failure (CHF) includes HFrEF and HFpEF. The binary classification was made as the presence or absence of CHF [27]. In the study in [28], the patient class in the data set included HFrEF, HFpEF, and Coronary Artery Disease. However, the classification was made as the presence or absence of HF, and the patient class was also classified according to the New York Heart Association (NYHA) functional classification. The focuses of the studies in [23], [24], [25], [26], [27], and [28] were not on the diagnosis of HFrEF and HFpEF.

A 10-year HF risk estimation was made for HFrEF and HFpEF with an AI-based system developed using 12-lead ECG data. Here, LVEF < 50% for HFrEF and LVEF \geq 50% for HFpEF were accepted. That was, mid-range ejection fraction (HFmrEF) ($41\% \leq$ LVEF \leq 49%) was also included in the HFrEF. There was no evidence this system could help distinguish between HFpEF and HFrEF [29].

HFrEF cases are easier to diagnose. HFrEF is diagnosed when LVEF \leq 40% is detected by echocardiography. The underlying phenotypic heterogeneity in HFpEF is more complex than in HFrEF [30]. The diagnosis of HFpEF is difficult even for specialist physicians. Because of some similar symptoms, cases of HF can be confused with cases of chest diseases. Since LVEF appears to be normal in the case of HFpEF, a patient who goes to the hospital with the symptoms of HF can be referred to a pulmonologist. When the functions of the patient evaluated in terms of chest diseases are normal, no action is taken. Therefore, the case of HFpEF is overlooked. If the pulmonologist is careful and considers the possibility of HFpEF and orders BNP and NT-proBNP blood tests, the case of HFpEF will be detected. (Brain natriuretic peptide (BNP) and N-terminal pro b-type natriuretic peptide (NT-proBNP) blood tests are often used to diagnose HF). To prevent such confusion and cases from being overlooked, it was aimed to design a medical decision support system using 3-lead ECG measured simply, and giving results within seconds. Thus, before performing echocardiography or other troublesome blood tests, a system that provides the doctor with preliminary information about the individual was HFrEF, HFpEF, or healthy was designed. It was thought that the system, advantageous both in terms of patient comfort and economically, would support situations that even specialist physicians have difficulty in diagnosing.

Only features extracted from 3-lead ECG data were used as features in the classification. To investigate whether only these features could be used without the need for demographic or echocardiographic data and other detailed examinations for the medical decision support system, the difference between classes was examined by statistical methods.

TABLE 1. Demographic information.

| Groups | Age (years) | Weight (kg) | Height (cm) | BMI (kg/m ²) |
|-------------------------------|---------------|---------------|---------------|--------------------------|
| Mean ± SD (range) for males | 65.18 ± 12.13 | 78.39 ± 13.22 | 170.72 ± 6.07 | 26.88 ± 4.35 |
| Mean ± SD (range) for females | 67.78 ± 15.67 | 75 ± 22.21 | 157.71 ± 7.08 | 39.26 ± 13.07 |
| Mean ± SD (range) for total | 66.37 ± 13.80 | 76.83 ± 17.83 | 164.75 ± 9.21 | 34.72 ± 12.26 |

Statistically the most distinctive features were determined, and classification studies were carried out with ML algorithms by reducing the number of features. The classification was carried out with high accuracy with the k-NN method.

II. MATERIAL AND METHODS

The data were collected from patients who visited Sakarya University Training and Research Hospital’s cardiology outpatient clinic as well as those who were admitted to the inpatient service. To carry out the study, the ethics committee report numbered 16214662/050.01.04/123 was obtained from the Sakarya University Faculty of Medicine Dean’s Office. An Informed Consent from each volunteer was obtained. Data collecting and recording were made with Biopac MP36 device. Data were obtained from individuals whose echocardiographic results were interpreted by a specialist cardiologist and were eligible for inclusion in our study.

ECG signals from 61 volunteers (25 years old or older) were taken from the right ankle and right and left wrists (Standard Bipolar Lead I). The sampling frequency of the signal was 200 Hz. Recordings were taken from each individual for 10 s. Data from the same volunteer at different time intervals were also used. The dataset contained a total of 180 data, of which 60 were HFrEF, 60 were HFpEF, and 60 were healthy individuals with LVEF above 50%. The demographic information of the volunteers was given in Table 1.

A. PREPROCESSING THE DATA

The workflow diagram of the study was provided in Figure 1. The study was shaped according to this flow diagram, and the results were obtained. After data were acquired from the volunteers, the data were labeled by a specialist cardiologist. First, noise and artifacts in the ECG signal were cleaned. For this, a Chebyshev Type II band-pass filter was applied in the range of 0.25 – 100 Hz, and Stopband Attenuation was 60 dB. A notch filter, in the range of 49-51 Hz, was applied to clean the mains noise at 50 Hz. Stopband Attenuation was also specified as 60 dB. Finally, the Moving Average Filter was applied to the signal.

A graphical representation of the ECGs of individuals with HFrEF, HFpEF, and healthy (LVEF above 50%) was given in Figure 2. In addition, the periodogram graph with the Fast Fourier Transforms of the ECGs was also given in the figure. Although the ECGs were very similar to each other, they were differentiated in the periodogram. Therefore, using a

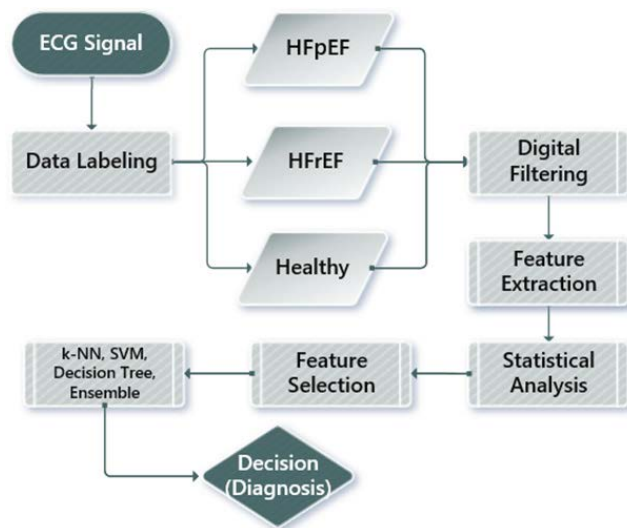


FIGURE 1. Flow diagram of study.

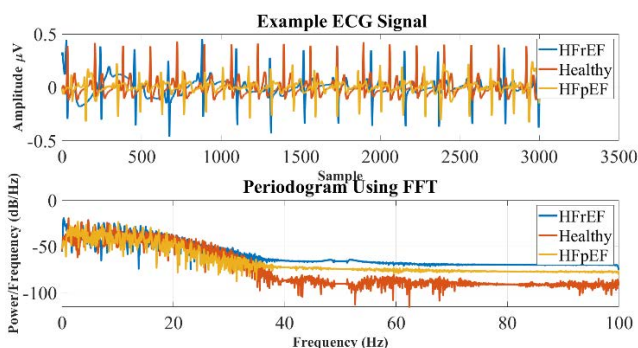


FIGURE 2. Periodogram Graph of ECG.

periodogram is a smart way for machine learning methods. Then, feature extraction from the ECG was performed. Statistical analysis was performed and features were selected. Finally, classification was made with the selected features and the diagnosis algorithm was developed.

B. ECG FEATURES

First, 21 features were extracted from the ECG in the time domain. These extracted features were given in Table 2 [31]. In the table, the numbers given to the features are in the first column, the names of the features are in the second column and the formulas of the features are in the third column. In addition to these features, 8 the Yule-Walker output

TABLE 2. ECG features at time domain.

| No | Feature | Equation |
|----|-------------------------------|--|
| 1 | Kurtosis | $x_{kur} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(n-1)S^4}$ |
| 2 | Skewness | $x_{ske} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)S^3}$ |
| 3 | * Interquartile range | IQR = iqr(x) |
| 4 | Geometric Mean | $G = (x_1 + \dots + x_n)^{1/n}$ |
| 5 | Harmonic Mean | $H = n / (\frac{1}{x_1} + \dots + \frac{1}{x_n})$ |
| 6 | Mobility – Hjort Parameters | $M = S_1^2 / S^2$ |
| 7 | Complexity – Hjort Parameters | $C = ((S_2^2 / S_1^2)^2 - (S_1^2 / S^2)^2)^{1/2}$ |
| 8 | Activity – Hjort Parameters | $A = S^2$ |
| 9 | *Maximum | $x_{max} = \max(x_i)$ |
| 10 | Median | $\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & : x \text{ odd} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & : x \text{ even} \end{cases}$ |
| 11 | *Mean Absolute Deviation | MAD = mad(x) |
| 12 | *Minimum | $x_{min} = \min(x_i)$ |
| 13 | *Central Moments | CM = moment(x, 10) |
| 14 | Mean | $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ |
| 15 | Average Curve Length | $CL = \frac{1}{n} \sum_{i=2}^n x_i - x_{i-1} $ |
| 16 | Standard Deviation | $S = (\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2)^{1/2}$ |
| 17 | Standard Error | $S_{\bar{x}} = S / \sqrt{n}$ |
| 18 | Shape Factor | $SF = x_{rms} / (\frac{1}{n} \sum_{i=1}^n \sqrt{ x_i })$ |
| 19 | *Singular Value Decomposition | SVD = svd(x) |
| 20 | *25% Trimmed Mean | TM25 = trimmean(x, 25) |
| 21 | *50% Trimmed Mean | TM50 = trimmean(x, 50) |

*The feature was computed using MATLAB
 \bar{x} : the mean of the distribution
 n : the number of observations of the sample
 S_1 : the standard deviation of the first derivative of the signal
 S_2 : the standard deviation of the second derivative of the signal
 x_{rms} : Root Mean Squared $x_{rms} = (\frac{1}{n} \sum_{i=1}^n |x_i|^2)^{1/2}$

parameters (4 of the normalized autoregressive (AR) parameters corresponding to the model of order 4, the estimated variance of 1 white noise input and 3 reflection coefficients) and 8 Burg’s Method output parameters (4 of the normalized autoregressive (AR) parameters corresponding to the model of order 4, 1 the estimated variance of the white noise input and 3 of the reflection coefficients) 16 features were also added to the features. A total of 37 features were extracted. These features were also calculated using MATLAB.

C. STATISTICAL ANALYSIS

The ECG signal is not normally distributed. So, it can be statistically analyzed with non-parametric test methods. Kruskal Wallis Test is preferred as the data are non-normally distributed and there are three or multiple classes. The test is the nonparametric version of the one-way ANOVA [32]. The aim of using this test was to investigate whether there is a significant difference between the classes (HFpEF, HFpEF, and Healthy).

TABLE 3. p values of features as a result of Kruskal Wallis test.

| FEATURE NO | P VALUE | FEATURE NO | P VALUE |
|------------|---------|------------|---------|
| 1 | 0.8337 | 20 | 0.0000 |
| 2 | 0.0000 | 21 | 0.0000 |
| 3 | 0.1095 | 22 | 0.6501 |
| 4 | 0.0847 | 23 | 0.7135 |
| 5 | 0.8753 | 24 | 0.9343 |
| 6 | 0.0007 | 25 | 0.9771 |
| 7 | 0.0000 | 26 | 0.1214 |
| 8 | 0.1583 | 27 | 0.0009 |
| 9 | 0.0014 | 28 | 0.0000 |
| 10 | 0.0000 | 29 | 0.1137 |
| 11 | 0.1784 | 30 | 0.0000 |
| 12 | 0.0107 | 31 | 0.0000 |
| 13 | 0.0464 | 32 | 0.0002 |
| 14 | 0.7619 | 33 | 0.1116 |
| 15 | 0.0261 | 34 | 0.0000 |
| 16 | 0.1583 | 35 | 0.0007 |
| 17 | 0.1583 | 36 | 0.0000 |
| 18 | 0.1457 | 37 | 0.0000 |
| 19 | 0.1574 | | |

TABLE 4. HFpEF - healthy binary Mann Whitney-U test results.

| FEATURE NO | R (ETA) | R ² | P-VALUE |
|------------|---------|----------------|---------|
| 2 | 0.2932 | 0.0860 | 0.0004 |
| 36 | 0.2136 | 0.0456 | 0.0029 |
| 12 | 0.1960 | 0.0384 | 0.0389 |
| 7 | 0.1835 | 0.0337 | 0.0333 |
| 20 | 0.1781 | 0.0317 | 0.0035 |
| 21 | 0.1758 | 0.0309 | 0.0131 |
| 9 | 0.1734 | 0.0301 | 0.9686 |
| 6 | 0.1580 | 0.0250 | 0.1482 |
| 10 | 0.1527 | 0.0233 | 0.0269 |
| 35 | 0.1482 | 0.0220 | 0.1497 |
| 27 | 0.1468 | 0.0215 | 0.1497 |
| 30 | 0.1320 | 0.0174 | 0.0765 |
| 28 | 0.1294 | 0.0168 | 0.0967 |
| 31 | 0.1014 | 0.0103 | 0.2345 |
| 13 | 0.0918 | 0.0084 | 0.5906 |
| 15 | 0.0907 | 0.0082 | 0.4235 |
| 32 | 0.0642 | 0.0041 | 0.4174 |
| 37 | 0.0524 | 0.0027 | 0.7548 |
| 34 | 0.0518 | 0.0027 | 0.4265 |

Asymptotic Significances (p values), obtained for each feature as a result of the Kruskal Wallis test, were provided in Table 3. According to the table, $p < 0.05$ for features 2, 6, 7, 9, 10, 12, 13, 15, 20, 21, 27, 28, 30, 31, 32, 34, 35, 36, 37. That was, these features were distinctive for at least two of the three classes being compared.

The Mann Whitney U test is a non-parametric test used to determine whether the two sampled groups are from the same population [33]. As post hoc tests, binary Mann Whitney-U tests were performed to determine whether the features that appeared distinctive as a result of the Kruskal Wallis test were distinctive only for certain binary classes or for all three classes. Mann-Whitney U test results were provided in Tables 4, 5, and 6. Features were ranked at these tables.

TABLE 5. HFrEF - HFpEF binary Mann Whitney-U test results.

| FEATURE NO | R (ETA) | R ² | P-VALUE |
|------------|---------|----------------|---------|
| 37 | 0.5576 | 0.3109 | 0.0000 |
| 2 | 0.3790 | 0.1437 | 0.0000 |
| 31 | 0.3381 | 0.1143 | 0.0002 |
| 30 | 0.3347 | 0.1121 | 0.0000 |
| 35 | 0.3280 | 0.1076 | 0.0002 |
| 27 | 0.3233 | 0.1045 | 0.0002 |
| 32 | 0.3144 | 0.0989 | 0.0053 |
| 6 | 0.2974 | 0.0884 | 0.0002 |
| 28 | 0.2668 | 0.0712 | 0.0056 |
| 34 | 0.1362 | 0.0186 | 0.0001 |
| 36 | 0.2350 | 0.0552 | 0.0235 |
| 7 | 0.1986 | 0.0395 | 0.0235 |
| 12 | 0.1218 | 0.0148 | 0.1976 |
| 9 | 0.0860 | 0.0074 | 0.0199 |
| 15 | 0.0839 | 0.0070 | 0.0420 |
| 10 | 0.0666 | 0.0044 | 0.0765 |
| 13 | 0.0405 | 0.0016 | 0.0837 |
| 20 | 0.0278 | 0.0008 | 0.0208 |
| 21 | 0.0236 | 0.0006 | 0.0464 |

TABLE 6. HFpEF - healthy binary Mann Whitney-U test results.

| FEATURE NO | R (ETA) | R ² | P-VALUE |
|------------|---------|----------------|---------|
| 2 | 0.6123 | 0.3750 | 0.0000 |
| 37 | 0.5713 | 0.3264 | 0.0000 |
| 30 | 0.5244 | 0.2750 | 0.0000 |
| 36 | 0.5075 | 0.2575 | 0.0000 |
| 31 | 0.4578 | 0.2096 | 0.0000 |
| 7 | 0.4171 | 0.1740 | 0.0000 |
| 28 | 0.3995 | 0.1596 | 0.0000 |
| 34 | 0.3870 | 0.1497 | 0.0000 |
| 32 | 0.3581 | 0.1283 | 0.0000 |
| 10 | 0.2605 | 0.0679 | 0.0000 |
| 21 | 0.1885 | 0.0355 | 0.0000 |
| 12 | 0.1799 | 0.0324 | 0.0055 |
| 20 | 0.0946 | 0.0089 | 0.0000 |
| 13 | 0.0917 | 0.0084 | 0.0108 |
| 9 | 0.0552 | 0.0030 | 0.0001 |
| 15 | 0.0242 | 0.0006 | 0.0131 |
| 6 | 0.1555 | 0.0242 | 0.0181 |
| 35 | 0.1718 | 0.0295 | 0.0202 |
| 27 | 0.1690 | 0.0286 | 0.0242 |

Since there is own error percentage for each of the classes, the value 0.05 is divided by 3 (= 0.0167) (This is named the Bonferroni correction.) [30]. When p-value is lower than 0.0167 as a result of the Mann Whitney U test for the relevant feature, the feature is said that it is distinctive for those two classes. That is, for example, features 2, 7, 9, 10, 12, 13, 15, 20, 21, 28, 30, 31, 32, 34, 36, and 37 were distinctive for HFpEF – Healthy binary classification because of $p < 0.0167$ for them.

Then, a different number of features from the most relevant ones were taken as input, and experiments were conducted. In the relationship of fewer features/higher accuracy, the optimum value was determined. Consequently, classification was performed with high accuracy as a result of the algorithm using features 2nd, 32nd, 36th, and 37th. These features

ranked highly in the Mann Whitney U test ranking. Triple classification average performance parameters were calculated [35] and provided in Table 7.

D. CLASSIFICATION

The dataset was organized as individuals in rows and features in columns. With tagging, a tag column was added to the end of the feature columns. The data were separated as 80% training and 20% testing (validation). Balanced test data were created by randomly taking 20% of the data in each class. Performance parameters were calculated by comparing the results obtained from the simulation with the predetermined label column.

Four different machine learning algorithms were applied as k-NN, Support Vector Machine (SVM), Decision Trees, and Ensemble Bagged Trees.

The k-NN classifier was fine k-NN [36]. The euclidean metric distance was used for the fine k-NN. Cubic SVM was applied [37]. The Kernel Function was polynomial, the Polynomial Order was 3, the Kernel Scale was automatic and the Box Constraint was 1. The applied Fine Decision Tree [38] Split criterion was Gini’s diversity index, the maximum number of splits was 100, and the Surrogate decision splits were off. Ensemble Bagged Trees, the ensemble classifier, was ensembled with the Bag Method [39], learner type was Decision Tree. The maximum number of splits was 179 and the Number of learners was 30.

Performance parameters for triple classification made with each classifier were provided in Table 7. The best results were obtained by the k-NN algorithm. Parameters were calculated separately for each of the three classes. For each parameter, the final value indicated in the table is the mean value of them.

The parameters were calculated using the following formulas [35]:

| | | PREDICTED Class | | |
|--------------|---------|-----------------|-------|---------|
| | | Classes | HFrEF | Healthy |
| ACTUAL Class | HFrEF | TN | FP | TN |
| | Healthy | FN | TP | FN |
| | HFpEF | TN | FP | TN |

FIGURE 3. Representation of Confusion Matrix.

For class i:

$$\text{Sensitivity}_i = \frac{TP_i}{TP_i + \sum FN_i}$$

$$\text{Specificity}_i = \frac{\sum TN_i}{\sum TN_i + \sum FP_i}$$

$$F_Measure = \frac{\sum_{i=1}^3 TP_i}{\text{Grand Total}}$$

$$\text{Kappa} = \frac{cs - \sum_{i=1}^3 p_i t_i}{s^2 - \sum_{i=1}^3 p_i t_i}$$

$c = \sum_{i=1}^3 C_{ii}$ the total number of elements correctly predicted

TABLE 7. Classification performance parameters.

| MACHINE LEARNING ALGORITHM | ACCURACY (%) | SENSITIVITY | SPECIFICITY | F-MEASURE | KAPPA | AUC |
|----------------------------|--------------|-------------|-------------|-----------|--------|--------|
| K-NN | 100 | 1 | 1 | 1 | 1 | 1 |
| SVMS | 97.22 | 0.9722 | 0.9861 | 0.9722 | 0.9583 | 0.9942 |
| DECISION TREES | 91.67 | 0.9167 | 0.9583 | 0.9167 | 0.8750 | 0.9780 |
| ENSEMBLE BAGGED TREES | 97.22 | 0.9722 | 0.9861 | 0.9722 | 0.9583 | 0.9896 |

$s = \sum_i^3 \sum_j^3 C_{ij}$ the total number of elements

$p_i = \sum_k^3 C_{ik}$ the number of times that class i was predicted (column total)

$t_i = \sum_k^3 C_{ki}$ the number of times that class i truly occurs (row total).

III. RESULTS AND DISCUSSION

In this study, a new diagnostic algorithm for HFrEF and HFpEF was developed. A machine learning based system was developed using features derived from ECG. A total of 37 features were derived. However, it is difficult to derive so many features in real-time systems. Therefore, it was aimed to reduce the number of features and improve the system. To increase the performance of the classifier, the 37 features were selected by the Kruskal Wallis Test and Mann Whitney-U Test methods and used.

The Kruskal Wallis test determined which of the features were distinctive for the triple classification (Table 3). As a result of this test, the number of features was reduced from 37 to 19. Then, these features were evaluated by the Mann Whitney-U Test for different binary groups, and ranked (Tables 4, 5, 6).

After that, a different number of features from the most relevant ones were taken as input, and experiments were carried out. The optimum value was determined in the fewer feature/high accuracy relationship. As a result, classification with high accuracy was performed as a result of the algorithm using the 2nd, 32nd, 36th, and 37th features. These features ranked highly in the Mann Whitney U test ranking. The classification performance parameters are provided in Table 7. The results in the table were obtained with only four features. This may be argued as an impressive performance for a system that may be used in practice.

Patients get various diagnostic tests, invasive processes and therapies during their illness and generate big amounts of data that can be collected in registries or other institutional databases to assess healthcare utilization, quality and cost of care, and prognosis [40]. For traditional analytical methods, the size, dynamic nature and complexity of this 'big data' can be demanding to make sense of [41]. ML methods can handle temporary, large volume and multimodal data [42]. ML methods are well-equipped for handling high-dimensional datasets with numerous variables, that complicates traditional statistical approaches such as regression. ML can also process collinear or correlated data points and evaluate complex interactions between predictors. ML algorithms enable accurate, higher performing

computation of nonlinear relationships [43]. ML comprises computational techniques that can take out patterns from data, get knowledge, and implement that knowledge to tasks such as risk prediction [44]. Providing the right care to the patient in HF is challenged by diagnostic ambiguity, variability in treatment and complexity in risk stratification, and restricted integration of information about care. ML can act an important role in filling these gaps in HF and has significant advantages over traditional human-induced models [45]. For these reasons, in this study, ML methods were preferred for classification. This study was thought to be applicable to larger datasets in the future and could be integrated into measurement devices.

Classifier results with four different ML algorithms were given in detail for each classifier in Table 8. The classification processes were carried out in order. First, 37 ECG features were classified without applying any feature selection algorithm, and performance parameters were calculated to measure the performance of the classifier and recorded in the relevant column. Then, 37 features were reduced to 19 with the first feature selection stage (Kruskal Wallis Test) and the same process was repeated. 19 features were reduced to four with the second feature selection stage (Mann Whitney-U Test) and the same process was repeated. The performance parameters for the classifiers were calculated and provided in the table.

According to the table, the best results were obtained with the k-NN classifier. At the table, reducing the number of features with feature selection generally did not decrease the system performance, but increased it even more. For k-NN, SVM, and Ensemble classifiers, this increase was quite evident. In addition, it was important to obtain the highest performance parameters with 4 features instead of 37 or 19 features, for real-time systems having less workload and for working more effectively. Although there was no significant increase in the performance parameters with feature selection for the Decision Trees classifier, in detail, the AUC increased for each class. It was also advantageous to obtain the performance parameters that could be obtained with 37 or 19 features, with 4 features.

Receiver operating characteristic (ROC) curves for the classifiers were provided in Figure 4. A ROC curve may be evaluated as: it can better diagnose the positive class if the curve is closer to the left axis; If the curve is closer to the upper axis, it can better define the control group. Regarding the figure, for example, in the ROC curve for the Healthy class, the positive class is "Healthy" and the control class is "HFrEF & HFpEF".

TABLE 8. Classifier results.

| k-NN Classifier Results | | | | | | | | | | | | |
|-----------------------------------|-----------------------------------|--------|---------|--------|-----------------------------------|--------|---------|--------|----------------------------------|--------|---------|--------|
| | Without Feature Selection (NF=37) | | | | After Kruskal Wallis Test (NF=19) | | | | After Mann Whitney-U Test (NF=4) | | | |
| | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF |
| Sensitivity | 0.9167 | 0.9167 | 0.9167 | 0.9167 | 0.9722 | 1 | 0.9167 | 1 | 1 | 1 | 1 | 1 |
| Specificity | 0.9583 | 0.9583 | 0.9583 | 0.9583 | 0.9861 | 0.9583 | 1 | 1 | 1 | 1 | 1 | 1 |
| AUC | 0.9375 | 0.9375 | 0.9375 | 0.9375 | 0.9792 | 0.9792 | 0.9583 | 1 | 1 | 1 | 1 | 1 |
| F-Measure | 0.9167 | | | | 0.9722 | | | | 1 | | | |
| Kappa | 0.8750 | | | | 0.9583 | | | | 1 | | | |
| Accuracy (%) | 91.67 | | | | 97.22 | | | | 100 | | | |
| SVMs Classifier Results | | | | | | | | | | | | |
| | Without Feature Selection (NF=37) | | | | After Kruskal Wallis Test (NF=19) | | | | After Mann Whitney-U Test (NF=4) | | | |
| | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF |
| Sensitivity | 0.8889 | 0.8333 | 0.8333 | 1 | 0.9167 | 0.8333 | 0.9167 | 1 | 0.9722 | 1 | 0.9167 | 1 |
| Specificity | 0.9444 | 0.9583 | 0.9583 | 0.9167 | 0.9583 | 1 | 0.9167 | 0.9583 | 0.9861 | 0.9583 | 1 | 1 |
| AUC | 0.9340 | 0.8785 | 0.9653 | 0.9583 | 0.9143 | 0.8333 | 0.9097 | 1 | 0.9942 | 1 | 0.9896 | 0.9931 |
| F-Measure | 0.8889 | | | | 0.9167 | | | | 0.9722 | | | |
| Kappa | 0.8333 | | | | 0.8750 | | | | 0.9583 | | | |
| Accuracy (%) | 88.99 | | | | 91.67 | | | | 97.22 | | | |
| Decision Trees Classifier Results | | | | | | | | | | | | |
| | Without Feature Selection (NF=37) | | | | After Kruskal Wallis Test (NF=19) | | | | After Mann Whitney-U Test (NF=4) | | | |
| | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF |
| Sensitivity | 0.9167 | 0.9167 | 0.9167 | 0.9167 | 0.9167 | 0.8333 | 1 | 0.9167 | 0.9167 | 0.9167 | 0.9167 | 0.9167 |
| Specificity | 0.9583 | 0.9167 | 0.9583 | 1 | 0.9583 | 0.9583 | 0.9167 | 1 | 0.9583 | 0.9167 | 0.9583 | 1 |
| AUC | 0.9537 | 0.9479 | 0.9774 | 0.9358 | 0.9352 | 0.8837 | 0.9514 | 0.9705 | 0.9780 | 0.9705 | 0.9878 | 0.9757 |
| F-Measure | 0.9167 | | | | 0.9167 | | | | 0.9167 | | | |
| Kappa | 0.8750 | | | | 0.8750 | | | | 0.8750 | | | |
| Accuracy (%) | 91.67 | | | | 91.67 | | | | 91.67 | | | |
| Ensemble Classifier Results | | | | | | | | | | | | |
| | Without Feature Selection (NF=37) | | | | After Kruskal Wallis Test (NF=19) | | | | After Mann Whitney-U Test (NF=4) | | | |
| | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF | TC | HFrEF | Healthy | HFpEF |
| Sensitivity | 0.9167 | 0.9167 | 0.8333 | 1 | 0.9167 | 0.8333 | 1 | 0.9167 | 0.9722 | 1 | 0.9167 | 1 |
| Specificity | 0.9583 | 0.9167 | 0.9583 | 1 | 0.9583 | 1 | 0.9167 | 0.9583 | 0.9861 | 0.9583 | 1 | 1 |
| AUC | 0.9699 | 0.9444 | 0.9653 | 1 | 0.9658 | 0.9375 | 0.9722 | 0.9878 | 0.9896 | 0.9826 | 0.9861 | 1 |
| F-Measure | 0.9167 | | | | 0.9167 | | | | 0.9722 | | | |
| Kappa | 0.8750 | | | | 0.8750 | | | | 0.9583 | | | |
| Accuracy (%) | 91.67 | | | | 91.67 | | | | 97.22 | | | |

NF number of features, TC Triple Classification

According to Figure 4, when there were 37 features for k-NN, the AUC for all three classes was 0.938. AUC values increased for all three classes when there were 19 features. HFpEF was ideal, the positive class was better discriminated for the Healthy class, while the control class (HFpEF & Healthy) was better discriminated for HFrEF. With 4 features, ideal results were achieved for all classes.

When there were 37 features for SVM, the control classes were better differentiated for HFpEF and Healthy classes, while the positive class was better differentiated for HFrEF. With 19 features, HFpEF was ideal, while AUC was slightly reduced for other classes. HFrEF was ideal when there were 4 features, and the situations were close to ideal for HFpEF and Healthy classes.

When there were 37 features for the Decision Tree, the positive class was better differentiated for HFpEF, while the control classes were better differentiated for the Healthy and HFrEF classes. The AUC value for HFrEF was significantly reduced when there were 19 features. Better results were obtained for all three classes when there were 4 features.

When there were 37 features for Ensemble Classifier, HFpEF was ideal, for HFrEF the control class was better

distinguished. When there were 19 features, the positive class for HFrEF was better differentiated. With 4 features, HFpEF was ideal, the control class was better differentiated for HFrEF, and the positive class was better differentiated for the Healthy class.

This study was considered to be implantable in measurement devices in the future. The k-NN classifier was focused on because it was widely used in industrial implementations [46]. Three other algorithms used were also used to support the study and to see the feasibility of this triple classification using 3-lead ECG. According to the results, this classification could be done with high accuracy and the best results were obtained with the k-NN method.

Compared to other studies in the field, there are three classes in the classification made in this study, namely HFrEF, Healthy, and HFpEF. Not only is the presence or absence of HF but also, if HF is present, the question “which one” is also answered.

This study contributed to the literature by showing that HFrEF and HFpEF could be diagnosed only with 3-lead ECG. It is thought that it will be a pioneer for future studies on this subject. The dataset used in the study was not very large.

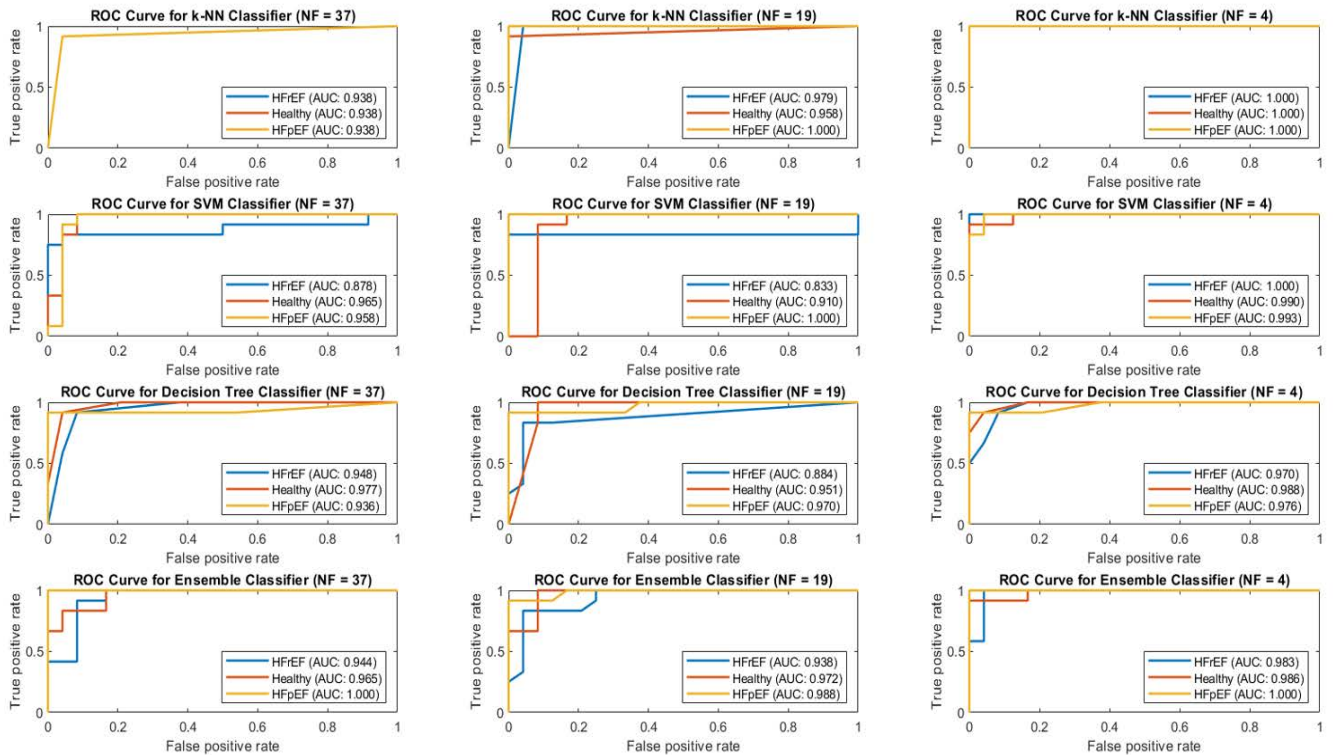


FIGURE 4. The ROC Curves for Without Feature Selection (NF = 37), After Kruskal Wallis Test (NF = 19), and After Mann Whitney-U Test (NF = 4) *NF* number of feature.

Therefore, traditional ML methods were used. New studies can be done using newer ML methods with a larger dataset.

IV. CONCLUSION

Echocardiogram and electrocardiogram (ECG) are the most frequently used tests in patients with suspected HF. LVEF is important not only as a manifestation of the disease but also because patients are defined by LVEF in most clinical studies. LVEF is generally determined by an echocardiogram. Rarely, SPECT (Single Photon Emission Computerized Tomography) and radionuclide ventriculography or radionuclide angiography (Multiple-gated acquisition - MUGA) are also used. These devices are pricey, some of the methods are invasive and all of them require an expert. In addition, there may be conditions where the attainment of these devices is limited. On the other hand, ECG is not expensive, easy to acquire, and gives fast results. Therefore, for example, when a patient admitted to the emergency service was suspected of HF, before carrying out detailed examinations, for informing the doctor whether the incoming case was HFpEF, HFrEF or healthy, a medical decision support system was developed using 3-lead ECG. Thus, diagnosis will be easier for the doctor, and time will be saved by preventing unnecessary and expensive tests. Therefore, it is a useful system both in terms of patient comfort and economics.

For HFrEF and HFpEF that was difficult to diagnose and can be overlooked, the medical decision support system that

did not require an expert and facilitated the diagnosis of the doctor was developed using only 3-lead ECG. It was thought that this study would fill the gap in the literature regarding simultaneous evaluating the diagnosis of HFrEF and HFpEF using only ECG.

REFERENCES

- [1] A. Maceira, S. Prasad, M. Khan, and D. Pennell, "Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance," *J. Cardiovascular Magn. Reson.*, vol. 8, no. 3, pp. 417–426, Jul. 2006, doi: [10.1080/10976640600572889](https://doi.org/10.1080/10976640600572889).
- [2] R. D. Dickstein, V. Dean, K. Dickstein, G. Filippatos, M. Tendera, P. Widimsky, and J. L. Zamorano, "ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008," *Eur. J. Heart Failure*, vol. 10, no. 10, pp. 933–989, Oct. 2008, doi: [10.1016/J.EJHEART.2008.08.005](https://doi.org/10.1016/J.EJHEART.2008.08.005).
- [3] W. J. Paulus, C. Tschöpe, J. E. Sanderson, C. Rusconi, F. A. Flachskampf, F. E. Rademakers, P. Marino, O. A. Smiseth, G. De Keulenaer, A. F. Leite-Moreira, A. Borbély, I. Édes, M. L. Handoko, S. Heymans, N. Pezzali, B. Pieske, K. Dickstein, A. G. Fraser, and D. L. Brutsaert, "How to diagnose diastolic heart failure: A consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the heart failure and echocardiography associations of the European society of cardiology," *Eur. Heart J.*, vol. 28, no. 20, pp. 2539–2550, Oct. 2007, doi: [10.1093/EURHEARTJ/EHM037](https://doi.org/10.1093/EURHEARTJ/EHM037).
- [4] B. A. Borlaug and W. J. Paulus, "Heart failure with preserved ejection fraction: Pathophysiology, diagnosis, and treatment," *Eur. Heart J.*, vol. 32, no. 6, pp. 670–679, Mar. 2011, doi: [10.1093/EURHEARTJ/EHQ426](https://doi.org/10.1093/EURHEARTJ/EHQ426).
- [5] S. F. Nagueh, K. J. Middleton, H. A. Kopelen, W. A. Zoghbi, and M. A. V. Quiñones, "Doppler tissue imaging: A noninvasive technique for evaluation of left ventricular relaxation and estimation of filling pressures," *J. Amer. College Cardiol.*, vol. 30, no. 6, pp. 1527–1533, Nov. 1997, doi: [10.1016/S0735-1097\(97\)00344-6](https://doi.org/10.1016/S0735-1097(97)00344-6).

- [6] S. F. Nagueh, C. P. Appleton, T. C. Gillebert, P. N. Marino, J. K. Oh, O. A. Smiseth, A. D. Waggoner, F. A. Flachskampf, P. A. Pellikka, and A. Evangelista, "Recommendations for the evaluation of left ventricular diastolic function by echocardiography," *J. Amer. Soc. Echocardiogr.*, vol. 22, no. 2, pp. 107–133, Feb. 2009, doi: [10.1016/J.ECHO.2008.11.023](https://doi.org/10.1016/J.ECHO.2008.11.023).
- [7] P. Ponikowski et al., "2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: The task force for the diagnosis and treatment of acute and chronic heart failure of the European society of cardiology (ESC) developed with the special contribution of the heart failure association (HFA) of the ESC," *Eur. Heart J.*, vol. 37, no. 27, pp. 2129–2200, Jul. 2016, doi: [10.1093/EURHEARTJ/EHW128](https://doi.org/10.1093/EURHEARTJ/EHW128).
- [8] D.-J. Choi, J. J. Park, T. Ali, and S. Lee, "Artificial intelligence for the diagnosis of heart failure," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–6, Dec. 2020, doi: [10.1038/s41746-020-0261-3](https://doi.org/10.1038/s41746-020-0261-3).
- [9] J. E. Ho, A. Lyass, D. S. Lee, R. S. Vasan, W. B. Kannel, M. G. Larson, and D. Levy, "Predictors of new-onset heart failure: Differences in preserved versus reduced ejection fraction," *Circulat., Heart Failure*, vol. 6, no. 2, pp. 279–286, Mar. 2013, doi: [10.1161/CIRCHEARTFAILURE.112.972828](https://doi.org/10.1161/CIRCHEARTFAILURE.112.972828).
- [10] B. Kolkusa, H. Hacilar, M. Kus, B. Bakir-Güngör, A. Aral, and V. Ç. Güngör, "Diagnosis of coronary heart disease via classification algorithms and a new feature selection methodology," *Int. J. Data Mining Sci.*, vol. 1, no. 1, pp. 8–15, 2019.
- [11] B. Nukala, "Heart disease classification comparison among patients and normal subjects using machine learning and artificial neural network techniques," *Int. J. Biosensors Bioelectron.*, vol. 7, no. 3, pp. 77–79, 2021, doi: [10.15406/IJBSBE.2021.07.00216](https://doi.org/10.15406/IJBSBE.2021.07.00216).
- [12] J.-M. Kwon, K.-H. Kim, H. J. Eisen, Y. Cho, K.-H. Jeon, S. Y. Lee, J. Park, and B.-H. Oh, "Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features," *Eur. Heart J.-Digit. Health*, vol. 2, no. 1, pp. 106–116, May 2021, doi: [10.1093/EHJDH/ZTAA015](https://doi.org/10.1093/EHJDH/ZTAA015).
- [13] G. B. Lim, "AI identifies distinct subgroups of HFpEF," *Nat. Rev. Cardiol.*, vol. 18, no. 6, p. 387, Mar. 2021, doi: [10.1038/s41569-021-00545-0](https://doi.org/10.1038/s41569-021-00545-0).
- [14] M. K. M. V. Ganesan, "Identification of systolic and diastolic heart failure progression with Krawtchouk moment feature-aided Harris hawks optimized support vector machine," *Signal, Image Video Process.*, vol. 16, no. 1, pp. 127–135, 2022, doi: [10.1007/S11760-021-01978-3](https://doi.org/10.1007/S11760-021-01978-3).
- [15] S. Selvaraj, P. L. Myhre, M. Vaduganathan, B. L. Claggett, K. Matsushita, D. W. Kitzman, B. A. Borlaug, A. M. Shah, and S. D. Solomon, "Application of diagnostic algorithms for heart failure with preserved ejection fraction to the community," *JACC, Heart Failure*, vol. 8, no. 8, pp. 640–653, Aug. 2020, doi: [10.1016/J.JCHF.2020.03.013](https://doi.org/10.1016/J.JCHF.2020.03.013).
- [16] F. Y. O. Abdalla, L. Wu, H. Ullah, G. Ren, A. Noor, H. Mkindu, and Y. Zhao, "Deep convolutional neural network application to classify the ECG arrhythmia," *Signal, Image Video Process.*, vol. 14, no. 7, pp. 1431–1439, 2020, doi: [10.1007/S11760-020-01688-2](https://doi.org/10.1007/S11760-020-01688-2).
- [17] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, S. J. Asirvatham, A. J. Deshmukh, B. J. Gersh, and R. E. Carter, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction," *Lancet*, vol. 394, no. 10201, pp. 861–867, 2019, doi: [10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0).
- [18] Z. I. Attia, S. Kapa, F. Lopez-Jimenez, P. M. McKie, D. J. Ladewig, G. Satam, P. A. Pellikka, M. Enriquez-Sarano, P. A. Noseworthy, T. M. Munger, S. J. Asirvatham, C. G. Scott, R. E. Carter, and P. A. Friedman, "Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram," *Nature Med.*, vol. 25, no. 1, pp. 70–74, Jan. 2019, doi: [10.1038/s41591-018-0240-2](https://doi.org/10.1038/s41591-018-0240-2).
- [19] J. Kwon, S. Y. Lee, K. Jeon, Y. Lee, K. Kim, J. Park, B. Oh, and M. Lee, "Deep learning-based algorithm for detecting aortic stenosis using electrocardiography," *J. Amer. Heart Assoc.*, vol. 9, no. 7, p. 14717, Apr. 2020, doi: [10.1161/JAHA.119.014717](https://doi.org/10.1161/JAHA.119.014717).
- [20] Z. Lin, Y. Gao, Y. Chen, Q. Ge, G. Mahara, and J. V. Zhang, "Automated detection of myocardial infarction using robust features extracted from 12-lead ECG," *Signal, Image Video Process.*, vol. 14, no. 5, pp. 857–865, 2020, doi: [10.1007/S11760-019-01617-Y](https://doi.org/10.1007/S11760-019-01617-Y).
- [21] A. Sbröllini, M. C. De Jongh, C. T. Haar, R. W. Treskes, S. Man, L. Burattini, and C. A. Swenne, "Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: A deep-learning approach," *Biomed. Eng. OnLine*, vol. 18, no. 15, pp. 1–17, 2019, doi: [10.1186/S12938-019-0630-9](https://doi.org/10.1186/S12938-019-0630-9).
- [22] Y. Zhang and M. Xia, "Application of deep neural network for congestive heart failure detection using ECG signals," *J. Phys., Conf. Ser.*, vol. 1642, no. 1, 2020, Art. no. 012021, doi: [10.1088/1742-6596/1642/1/012021](https://doi.org/10.1088/1742-6596/1642/1/012021).
- [23] A. García-Escobar, S. Vera-Vera, A. Jurado-Román, S. Jiménez-Valero, G. Galeote, and R. Moreno, "Subtle QRS changes are associated with reduced ejection fraction, diastolic dysfunction, and heart failure development and therapy responsiveness: Applications for artificial intelligence to ECG," *Ann. Noninvasive Electrocardiol.*, vol. 2022, Jul. 2022, Art. no. e12998, doi: [10.1111/ANEC.12998](https://doi.org/10.1111/ANEC.12998).
- [24] D. K. Plati, E. E. Tripoliti, G. S. Karanasiou, A. Rammos, A. Bechlioulis, C. J. Watson, and K. McDonald, "Machine learning techniques for predicting and managing heart failure," *Predicting Heart Failure, Invasive, Non-Invasive, Mach. Learn. Artif. Intell. Based Methods*, vol. 2022, pp. 189–226, Apr. 2022, doi: [10.1002/9781119813040.CH9](https://doi.org/10.1002/9781119813040.CH9).
- [25] X. Jin, J. F. Nauta, C.-L. Hung, W. Ouwerkerk, T.-H. K. Teng, A. A. Voors, C. S. P. Lam, and Joost P. van Melle, "Left atrial structure and function in heart failure with reduced (HFREF) versus preserved ejection fraction (HFpEF): Systematic review and meta-analysis," *Heart Fail. Rev.*, vol. 1, pp. 1–23, Jan. 2022, doi: [10.1007/S10074-021-10204-8](https://doi.org/10.1007/S10074-021-10204-8).
- [26] R. Herman, M. Vanderheyden, B. Vavrik, M. Beles, T. Palus, O. Nelis, M. Goethals, S. Verstreken, R. Dierckx, M. Penicka, W. Heggermont, and J. Bartunek, "Utilizing longitudinal data in assessing all-cause mortality in patients hospitalized with heart failure," *ESC Heart Failure*, vol. 2022, pp. 1–10, Jun. 2022, doi: [10.1002/EHF2.14011](https://doi.org/10.1002/EHF2.14011).
- [27] Z. Mašetić and A. Subasi, "Congestive heart failure detection using random forest classifier," *Comput. Methods Programs Biomed.*, vol. 130, pp. 54–64, Jul. 2016, doi: [10.1016/j.cmpb.2016.03.020](https://doi.org/10.1016/j.cmpb.2016.03.020).
- [28] E. Hasumi, K. Fujii, Y. Chen, Y. Shimizu, T. Oshima, H. Matsunaga, and J. Matsuda, "Heart failure grading using single-lead electrocardiography," *medRxiv*, vol. 2020, pp. 1–34, Jan. 2020, doi: [10.1101/2020.10.08.20209700](https://doi.org/10.1101/2020.10.08.20209700).
- [29] I. Karabayir, L. Butler, D. Kitzman, A. Alonso, G. Tison, L. Y. Chen, G. Clifford, E. Z. Soliman, and O. Akbilgic, "PO-703-04 ECG-AI can predict risk for heart failure with both preserved and reduced ejection fraction," *Heart Rhythm*, vol. 19, no. 5, p. S447, May 2022, doi: [10.1016/j.hrthm.2022.03.1058](https://doi.org/10.1016/j.hrthm.2022.03.1058).
- [30] A. B. Gevaert, R. Kataria, F. Zannad, A. J. Sauer, K. Damman, K. Sharma, S. J. Shah, and H. G. C. Van Spall, "Heart failure with preserved ejection fraction: Recent concepts in diagnosis, mechanisms and management," *Heart*, vol. 108, no. 17, pp. 1342–1350, Sep. 2022, doi: [10.1136/HEARTJNL-2021-319605](https://doi.org/10.1136/HEARTJNL-2021-319605).
- [31] M. R. Bozkurt, M. K. Uçar, F. Bozkurt, and C. Bilgin, "In obstructive sleep apnea patients, automatic determination of respiratory arrests by photoplethysmography signal and heart rate variability," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 4, pp. 959–979, Dec. 2019, doi: [10.1007/s13246-019-00796-9](https://doi.org/10.1007/s13246-019-00796-9).
- [32] P. E. McKight and J. Najab, "Kruskal-Wallis test," in *The Corsini Encyclopedia of Psychology*, I. B. Weiner and W. E. Craighead, Eds. 2010, doi: [10.1002/9780470479216.corpsy0491](https://doi.org/10.1002/9780470479216.corpsy0491).
- [33] P. E. McKight and J. Najab, "Mann-Whitney U test," in *The Corsini Encyclopedia of Psychology*, I. B. Weiner and W. E. Craighead, Eds. 2010, doi: [10.1002/9780470479216.corpsy0524](https://doi.org/10.1002/9780470479216.corpsy0524).
- [34] W. Haynes, "Bonferroni correction," in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K. H. Cho, and H. Yokota, Eds. New York, NY, USA: Springer, 2013, doi: [10.1007/978-1-4419-9863-7_1213](https://doi.org/10.1007/978-1-4419-9863-7_1213).
- [35] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," 2020, *arXiv:2008.05756*.
- [36] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, p. 218, Jun. 2016, doi: [10.21037/ATM.2016.03.37](https://doi.org/10.21037/ATM.2016.03.37).
- [37] S. Ekiz and P. Erdoğmuş, "Comparative study of heart disease classification," in *Proc. Electr. Electron., Comput. Sci., Biomed. Eng. Meeting (EBBT)*, Apr. 2017, pp. 1–4, doi: [10.1109/EBBT.2017.7956761](https://doi.org/10.1109/EBBT.2017.7956761).
- [38] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991, doi: [10.1109/21.97458](https://doi.org/10.1109/21.97458).
- [39] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996, doi: [10.1007/BF00058655](https://doi.org/10.1007/BF00058655).
- [40] J. A. Cook and G. S. Collins, "The rise of big clinical databases," *Brit. J. Surg.*, vol. 102, no. 2, pp. e93–e101, Jan. 2015, doi: [10.1002/BJS.9723](https://doi.org/10.1002/BJS.9723).
- [41] A. B. Docherty and N. I. Lone, "Exploiting big data for critical care research," *Current Opinion Crit. Care*, vol. 21, no. 5, pp. 467–472, Oct. 2015, doi: [10.1097/MCC.0000000000000228](https://doi.org/10.1097/MCC.0000000000000228).

- [42] Y. Demchenko, P. Grosso, C. De Laat, and P. V. Membrey, "Addressing big data issues in scientific data infrastructure," in *Proc. Int. Conf. Collaboration Technol. Syst. (CTS)*, May 2013, pp. 48–55, doi: [10.1109/CTS.2013.6567203](https://doi.org/10.1109/CTS.2013.6567203).
- [43] D. L. Weller, T. M. T. Love, and M. V. Wiedmann, "Interpretability versus accuracy: A comparison of machine learning models built using different algorithms, performance measures, and features to predict *E. coli* levels in agricultural water," *Frontiers Artif. Intell.*, vol. 4, p. 14, May 2021, doi: [10.3389/FRAI.2021.628441](https://doi.org/10.3389/FRAI.2021.628441).
- [44] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. V. P. Campbell, "Introduction to machine learning, neural networks, and deep learning," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 14, 2020. [Online]. Available: <https://tvst.arvojournals.org/article.aspx?articleid=2762344>
- [45] T. Averbuch, K. Sullivan, A. Sauer, M. A. Mamas, A. A. Voors, C. P. Gale, M. Metra, N. Ravindra, and H. G. C. Van Spall, "Applications of artificial intelligence and machine learning in heart failure," *Eur. Heart J.-Digit. Health*, vol. 3, no. 2, pp. 311–322, Jul. 2022, doi: [10.1093/ehjdh/ztac025](https://doi.org/10.1093/ehjdh/ztac025).
- [46] C. Kier and T. Aach, "Predicting the benefit of sample size extension in multiclass k-NN classification," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, 2006, pp. 332–335, doi: [10.1109/ICPR.2006.942](https://doi.org/10.1109/ICPR.2006.942).



PINAR ÖZGEN KAVAS (Member, IEEE) received the bachelor's degree in electrical and electronics engineering from Istanbul University and the master's degree from Dumlupınar University. She is currently pursuing the doctoral degree with Sakarya University. She is a Research Assistant with the Department of Computer Engineering, Kütahya Dumlupınar University. Her research interests include biomedical signals, machine learning, and artificial intelligence.



MEHMET RECEP BOZKURT was born in İstanbul, Türkiye, in 1976. He received the B.S., M.S., and Ph.D. degrees in electronics engineering from the Sakarya University, Türkiye, in 1997, 2001, and 2007, respectively. From 1999 to 2008, he was a Research Assistant at the Electrical and Electronics Engineering Department, Sakarya University, where he was an Assistant Professor, from 2008 to 2015, and he was an Associate Professor, from 2015 to 2021. He has been working as a Professor with the Electrical and Electronics Engineering Department, Sakarya University, since 2021. He is the author of nearly 100 articles, and he has supervised eight completed doctoral studies. His research interests include biomedical signals, biopotentials, machine learning, and artificial intelligence.



İBRAHİM KOCAYIĞIT received the graduate degree from the Faculty of Medicine, Pamukkale University, in 2008, and a cardiology residency from the Sakarya University Education and Research Hospital, in 2013. He is currently an Assistant Professor with the Department of Cardiology, Faculty of Medicine, University of Sakarya. His research interests include complex coronary interventions and peripheral artery diseases.



CAHİT BILGIN was born in Bartın, Turkey, in 1964. He received the bachelor's degree from the Medical Faculty, Ankara University, Ankara, Turkey, in 1988, and the M.D. degree in chest disease from Düzce Medical Faculty, Düzce, Turkey, in 2003. He studied English Preparatory from the Computer Engineering Department, Hacettepe University, in 1981. He was a Lecturer at the Düzce Medical Faculty, from 2004 to 2007. He worked Hendek Government Hospital Sleep Laboratory, Sakarya, Turkey, from 2007 to 2014. Currently, he is an Associate Professor with the Department of Chest Disease, Training and Research Hospital, Sakarya University, Sakarya. His clinical interests include sleep medicine and portable devices, and his research interest includes sleep laboratory.

...