Classification of Atrial Fibrillation and Acute Decompensated Heart Failure Using Smartphone Mechanocardiography: A Multi-label Learning Approach

Saeed Mehrang*, Olli Lahdenoja, Matti Kaisti, Mojtaba Jafari Tadi, Tero Hurnanen, Antti Airola, Timo Knuutila, Jussi Jaakkola, Samuli Jaakkola, Tuija Vasankari, Tuomas Kiviniemi, Juhani Airaksinen, Tero Koivisto, and Mikko Pänkäälä

***Abstract*—Timely diagnosis of cardiovascular diseases (CVD) is crucial to prevent morbidity and mortality. Atrial fibrillation (AFib) and heart failure (HF) are two prevalent cardiac disorders that are associated with a high risk of morbidity and mortality, especially if they are concurrently present. Current approaches fail to screen many at-risk individuals who would benefit from preventive treatment; while others receive unnecessary interventions. An effective approach to the detection of CVDs is mechanocardiography (MCG) by which translational and rotational precordial chest movements are monitored. In this study, we collected MCG data from a study sample of 300 hospitalized cardiac patients using multidimensional built-in inertial sensors of a smartphone. Our main objective was to detect concurrent AFib and acute decompensated HF (ADHF) using smartphone MCG (or sMCG). To this end, we adopted a supervised machine learning classification using multi-label and hierarchical classification. Logistic regression, random forest, and extreme gradient boosting were used as candidate classifiers. The results of the analysis showed the area under the receiver operating characteristic curve values of 0.98 and 0.85 for AFib and ADHF, respectively. The highest percentages of positive and negative predictive values for AFib were 91.9 and 100; while for ADHF, they were 56.9 and 88.4 for the multi-label classification and 69.9 and 68.8 for the hierarchical classification, respectively. We conclude that using a single sMCG measurement, AFib can be detected accurately whereas ADHF can be detected with moderate certainty.**

***Index Terms*—Acute decompensated heart failure, atrial fibrillation, gyrocardiography, machine learning, seismocardiography, smartphone mechanocardiography**

I. Introduction

ATRIAL fibrillation (AFib) and heart failure (HF) are two of the most prevalent cardiac disorders. They frequently occur simultaneously and are associated with an elevated risk of morbidity and mortality [1]. The association between AFib and HF is important as each disease entity may complicate the diagnosis and treatment of the other [2]. AFib is a common arrhythmia characterized by chaotic electrical activation of the atria leading to a rapid and irregular heartbeat which in turn impedes the mechanical functioning of the heart. HF refers to impaired cardiac function, and unlike AFib, HF has no single characteristic feature. The clinical characteristics of HF are more complex when compared to AFib; involving the presence of multiple symptoms and signs that occur simultaneously, such as orthopnea and abnormal heart sounds [3]. Chronic HF can be compensated or decompensated. In compensated HF, the symptoms are stable with no sign of fluid retention and pulmonary edema. In decompensated HF (ADHF) the symptoms can be acute, as in cases where acute episodes of pulmonary edema, reduction of exercise tolerance, and increasing breathlessness on exertion are present [4]. The cause or causes of ADHF may include recurrent ischemia, arrhythmias, infections, and electrolyte disturbances [4].

Today, many invasive and non-invasive cardiac condition assessment tests are used for the diagnosis of cardiovascular diseases. Electrocardiography (ECG) is a common technique and considered the gold standard for a non-invasive diagnosis of cardiac arrhythmia. While abnormal ECG increases the likelihood of HF, it still lacks sufficient specificity [5], [6], [3]. Other tests such as echocardiography, stress testing, cardiac catheterization, biochemical testing, chest X-ray, cardiac magnetic resonance imaging, and radionuclide ventriculography are used to detect HF. These tests, however, require advanced equipment and sophisticated medical interpretations which makes them inapplicable for day-to-day life or home monitoring purposes [7], [8].

Ballistocardiography (BCG) measures the ballistic forces induced by the heart. Recently, BCG has been shown effective in tracking the clinical status of HF patients [9]. Seismocardiography (SCG) is an accelerometer-based sensing method that measures precordial translational movements [10], [11]. SCG is known to have the potential for detecting the presence of multiple heart diseases [12], [13], [14]. It has been recently used to assess the status of HF patients by analyzing cardiac response to exercise [15]. We considered the use of another sensor, similar to accelerometer, a gyroscope, to measure precordial rotational vibrations. This new measurement technique, also known as gyrocardiography (GCG), can yield information on myocardial mechanical motions [16]. Jointly, SCG and GCG together constitute the concept of mechanocardiography

S. Mehrang, O. Lahdenoja, M. Kaisti, M. Jafari Tadi, T. Hurnanen, T. Koivisto, and M. Pänkäälä are with the Department of Future Technologies, University of Turku, Turku, Finland (e-mail: firstname.lastname@utu.fi).

J. Jaakkola, S. Jaakkola, T.Vasankari, T. Kiviniemi, and J. Airasksinen are with Heart Center, Turku University Hospital, Turku, Finland e-mail: (firstname.lastname@tyks.fi).

S. Mehrang and O. Lahdenoja contributed equally to the preparation of this manuscript.

(MCG). Using the built-in inertial sensors of smartphones, it is possible to collect MCG data which may help to detect heart arrhythmia, myocardial infarction, and coronary artery diseases [17], [18], [19], [20].

In this study, we considered a sample size of 300 individuals having either AFib, both AFib and ADHF, or neither of them. We investigated the ability of smartphone mechanocardiography (sMCG) to detect potentially concurrent AFib and ADHF by measuring MCG signals using a smartphone and deploying (1) a multi-label classification framework [21] as well as (2) a hierarchical classification framework. A multi-label classification framework was used as our intention was to provide the means to classify every subject into multiple categories using an identical feature vector. In contrast, when the hierarchical classification framework was used, the AFib classification was first performed and later the ADHF classification was executed. The rationale behind the hierarchical classification approach is that those individuals who have AFib are more likely to develop ADHF [1]. Therefore, we tended to restrict the search for ADHF to only the AFib patients. A schematic diagram of the implemented study procedure is presented in Fig. 1. To the best of our knowledge, this is the first time that sMCG signals have been used for the detection of a simultaneous presence of AFib and ADHF.

II. Methods

A. Study protocol

Data collection for this paper was accomplished by placing a smartphone on the chest of the patients while lying down in a supine position on a hospital bed. Although the measurements were made by a single study investigator, it is possible for patients to carry out the measurement themselves as the method does not require specialized equipment and training. The study protocol was approved by the Ethics Committee of the Hospital District of Southwest Finland (ClinicalTrials.gov: NCT03274583). Written informed consent was acquired from all participants. The Helsinki declaration was strictly followed throughout the study.

A group of cardiac patients (n=300) was recruited and studied at Turku University Hospital, Turku, Finland. Of the 300 participants in this study, there were 75 ADHF cases and 150 AFib cases. As depicted in Fig. 1, all of the ADHF cases also had AFib. This data set was originally collected for a study on AFib detection; therefore, not all types of ADHF are present in this data set. The demographics of the subjects are presented in Table I.

B. Data acquisition

An MCG recording of three minutes was acquired using a Sony Xperia either Z1 or Z5 smartphone running on the Android operating system. Each recording contains simultaneous data streams from both a triaxial accelerometer and a triaxial gyroscope (altogether six channels) sampled with 200 Hz. The anonymity of the subjects was assured by assigning an ID number to each subject inside the signal acquisition smartphone application.

The recording was performed by placing the smartphone longitudinally on the subject's bare chest with the screen facing upwards and the bottom edge of the phone at the level of the lower edge of the sternum. The subjects were asked to remain still and avoid speaking while measurement was in progress. Termination of the measurement was handled automatically in the signal acquisition smartphone application. In addition to the MCG data recorded by the smartphone, a continuous 5-lead telemetry ECG (Philips IntelliVue MX40) was also acquired. The ECG recordings were used as the benchmark for assessing cardiac rhythm and the number of possible supraventricular (SVES) or ventricular extrasystoles (VES). The rhythm of each telemetry ECG was classified as either non-AFib, AFib or other by two independent cardiologists. In cases where there was disagreement between the classification of the two cardiologists, a third independent cardiologist – blinded to the original rhythm classification – made the final decision [18]. The presence of HF and its type was later decided by a cardiologist who went through the background data of the participants, as well as their echocardiography, and B-type natriuretic peptide (BNP) test results.

The data was then transferred to the facilities of the Department of Future Technologies, University of Turku, Finland. Pre-processing, feature extraction, and classification were performed with Matlab and Python.

C. Data Analysis Pipeline

1) Signal pre-processing: The signal pre-processing started with filtering each of the six channels by a bandpass filter. A brick-wall FFT (Fast Fourier Transform) filter with a passband of 1-45 Hz was applied to the SCG and GCG signals, allowing the removal of baseline wander and high-frequency noise. These frequency bands were obtained empirically. Breathing removal was also carried out separately on each channel of the inertial sensors by subtracting the low-pass filtered signal (moving average filter of length 50 samples) from the bandpass filtered signal.

2) Machine learning: Two different machine learning approaches were adopted for this study, namely multi-label and hierarchical classification as described below.

There are many different multi-label classification approaches, each having their advantages and disadvantages. Depending on how the multi-label classification method takes advantage of the correlations between the classes, there are more computationally complex and more relaxed versions of the algorithms. Perhaps, the most well-known method is the binary relevance [22], [23], [24], [21], where the evaluation of the multi-label classifier with k binary classes is based on simply implementing the k two-class classifiers. The results of the binary relevance are then combined into a multi-label vector/array. In our study, the hierarchical classification approach [25] describes the case where we first classify AFib and then classify ADHF for those participants who had been classified as AFib. Detecting ADHF for those individuals who have AFib is crucial as the risk of developing ADHF increases if AFib is present.

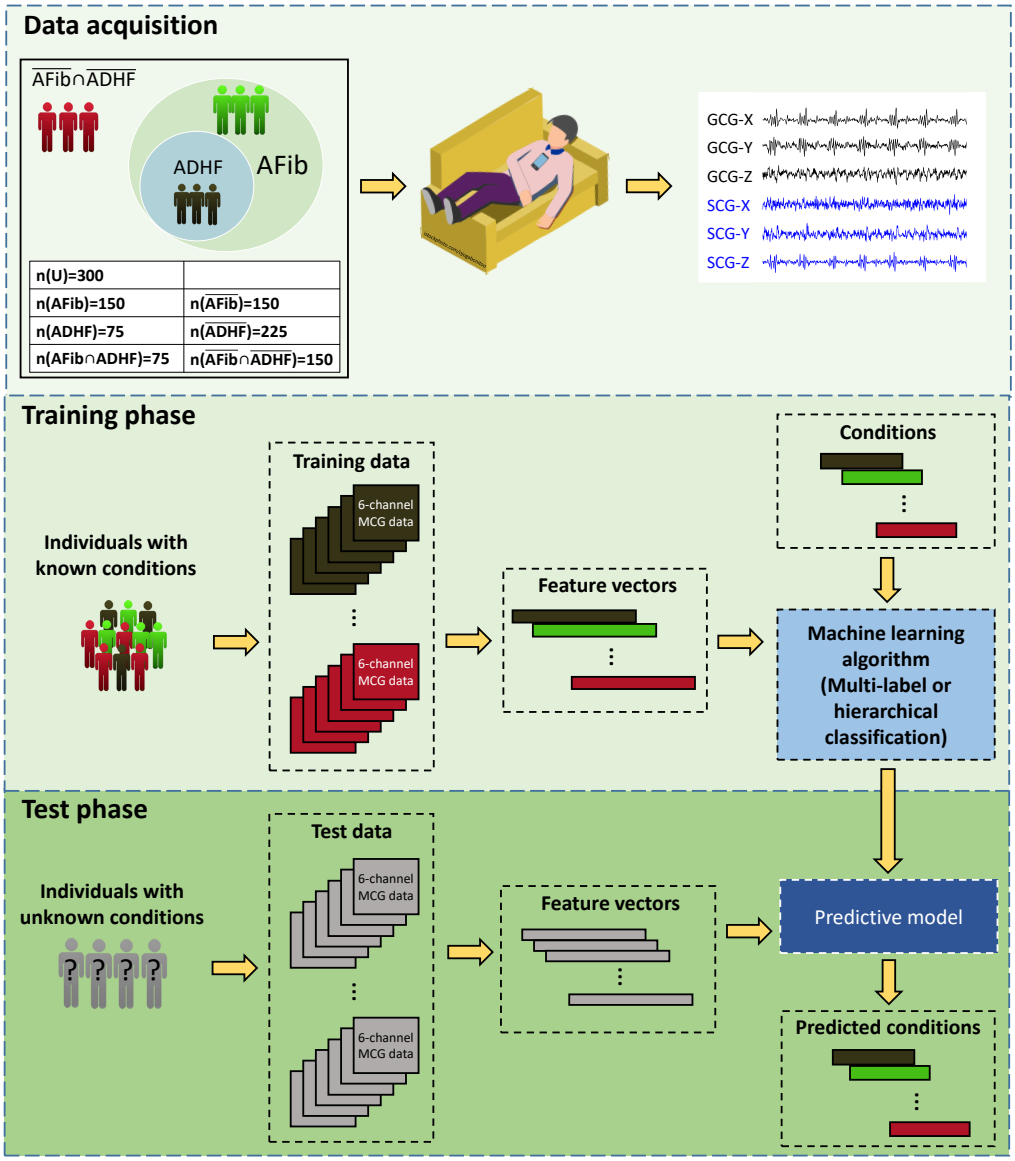


Fig. 1: Schematic diagram of the implemented research process in this study.

TABLE I: Demographics of the participants in the study. Mean and standard deviation of each variable are provided.

	Count	Age (years)	Gender (M)	Weight (kg)	Height (cm)	BMI (kg/m^2)
AFib	150	75.0 (9.7)	83	85.5 (19.4)	171.5 (10.0)	29.0 (5.7)
Non-AFib	150	74.5 (9.5)	85	78.7 (14.6)	169.2 (9.0)	27.5 (4.7)
ADHF	75	77.0 (7.9)	39	84.5 (18.5)	170.3 (9.7)	29.0 (5.6)
Non-ADHF	225	74.0 (9.9)	132	81.3 (17.1)	170.4 (9.6)	27.9 (5.1)

The feature matrix formation process is presented in Fig. 2 which contains signal pre-processing, signal segmentation, feature extraction, and feature matrix formation. First, the raw signals were pre-processed by bandpass filtering for noise removal. Next, the pre-processed signals were divided into non-overlapping segments of length 10 s [17]. There was a total of 5244 unique segments which translated into 17.5 segments (std = 1.3) per subject on average. Then, a feature vector was calculated using the data from each of the six channels of the SCG and GCG signals in each 10 s segment. The feature vector corresponding to each segment is a concatenation of the feature vectors derived from each of the 6 channels [17]. Next, for every subject, the feature vectors - from all 10 s segments - were placed into the rows of a matrix. The median value of each feature over all the segments was calculated afterwards. As a result, the whole feature matrix for each subject was

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2020.2981334, IEEE Sensors Journal

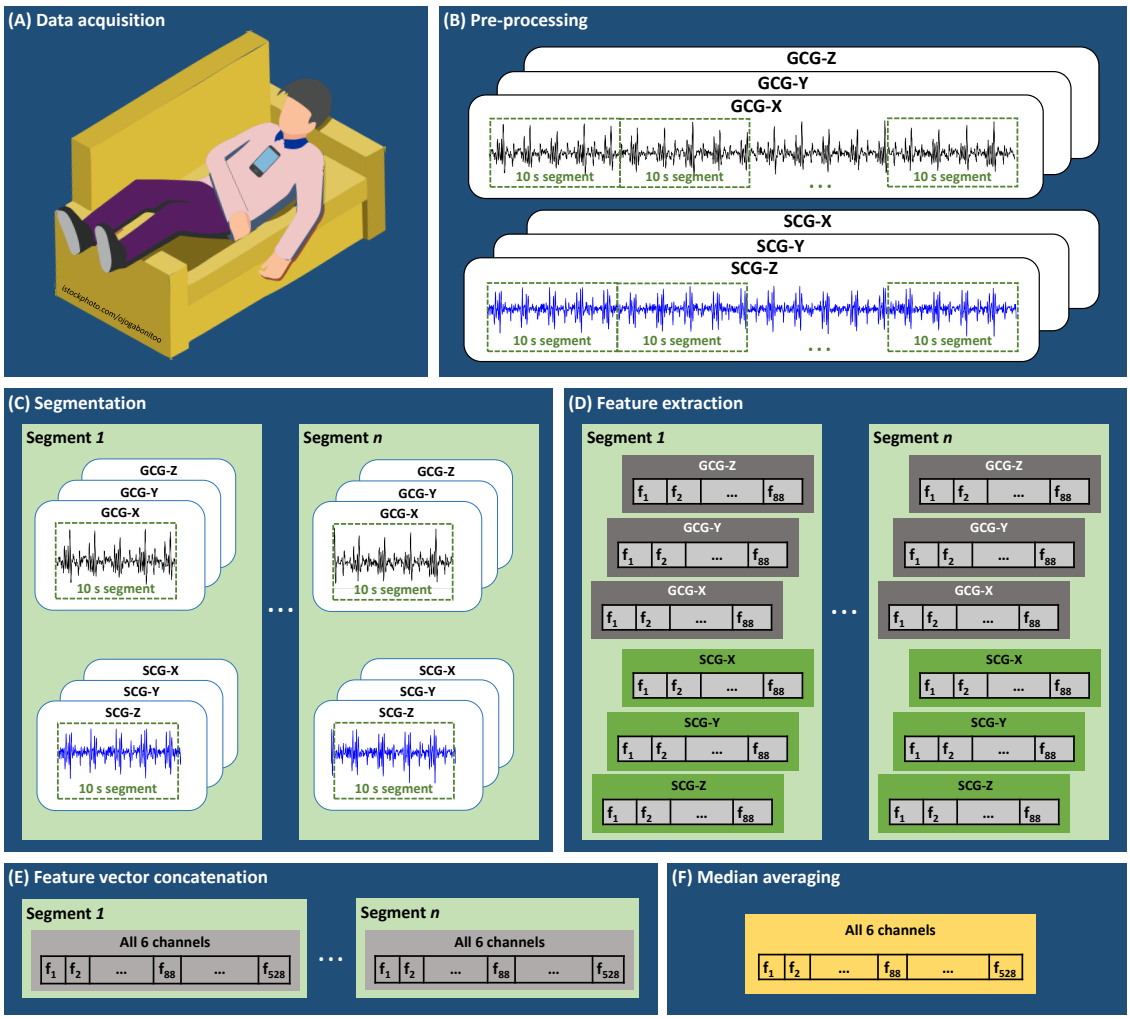


Fig. 2: Diagram of the feature matrix formation process. Following the flow of the diagram, there are six blocks from the top left to the bottom right. From the top left, the blocks represent (A) data acquisition, (B) pre-processing, (C) signal segmentation, (D) feature extraction (88 features (f) from each channel in each segment), (E) feature vector concatenation (one feature vector for each 6-channel segment comprising of 528 $(6 * 88)$ features), and (F) the final median averaging over the segments (one feature vector for the whole measurement). Note that, the plotted MCG signals correspond to a control participant.

abstracted into a single feature vector. The median averaging has shown better generalization in previous studies [26], [27]. Lastly, all abstracted feature vectors were placed into a feature matrix with a single row corresponding to each study subject.

Three different classifiers, namely random forest (RF) [28], extreme gradient boosting (XGB) [29], and logistic regression (LR) [30] were deployed for the multi-label classification task. Evaluation of the classifier was done using leave-one-subject-out cross-validation [31]. The cross-validation was nested [32] so that the inner cross-validation was used for hyper-parameter optimization and selection and the outer for testing the classifier with a fully unseen subject's data. In nested-cross validation, the classification process is done in a 7-step procedure. The procedure includes: (1) leaving out one subject's features from the feature matrix, (2) using the remaining feature set (F), optimizing the classifier parameters deploying stratified K-fold cross-validation [33] and randomized grid search [34] over the parameter space, (3) training the classifier using the identified set of optimal parameters and F, (4) predicting the label of the left-out subject (unseen subject), (5) storing the predicted label and the true label of the left-out subject, (6) placing the left-out subject's features back into the feature matrix, and (7) repeating steps 1 to 6 for all subjects. Please see Fig. 3 for a visual representation of the nested cross-validation in the multi-label and hierarchical classification approaches.

The classification tasks were all performed with Scikit-learn package in Python [34].

3) Feature extraction: Heart rate (HR) and heart rate variability (HRV), can be of relative help in distinguishing the regular or irregular patterns of the heart rhythm (e.g. in the case of AFib) as well as cardio-pulmonary condition assessment [35]. However, there are other major determinants of HR irregularity which need to be considered. Hence, to

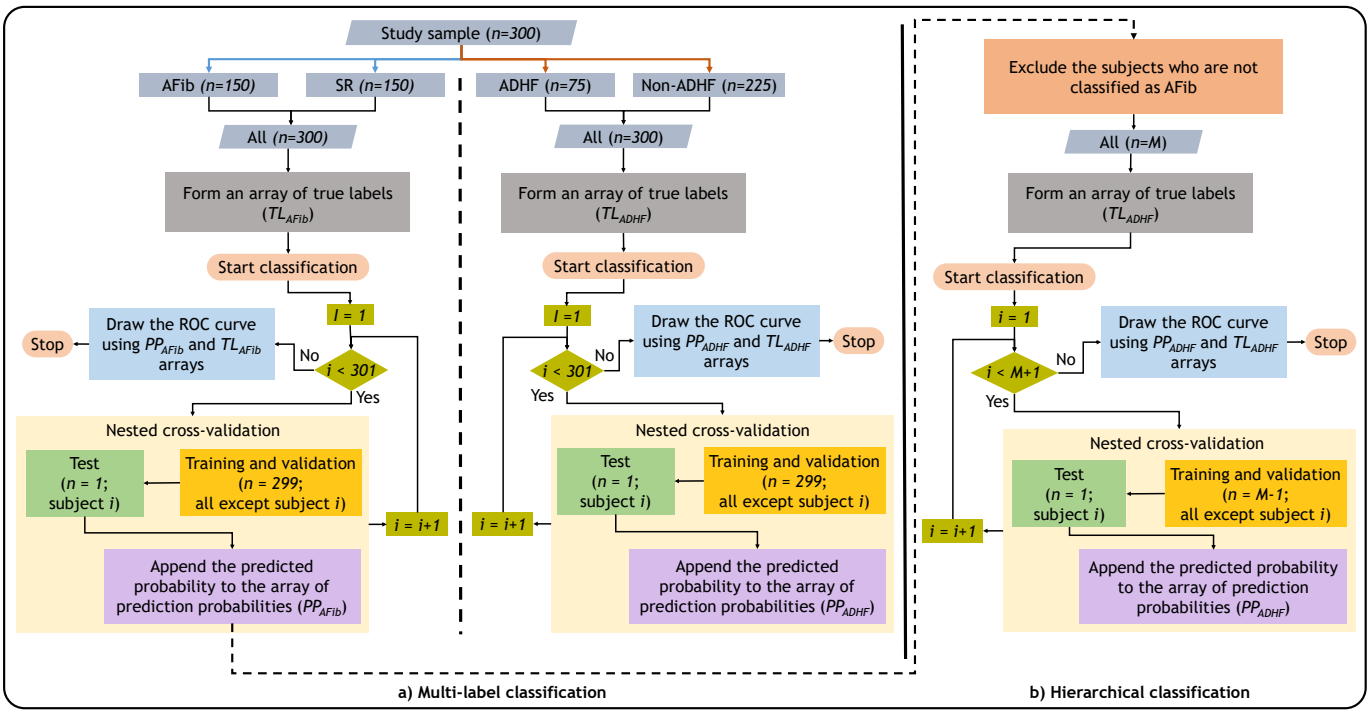


Fig. 3: Flowcharts of the nested cross-validation implemented for (a) multi-label and (b) hierarchical classification approaches.

obtain better discrimination accuracy, we considered the use of other contributors such as a self-similarity index and a distribution of spectral content within the signal by calculating approximate and spectral entropies, respectively. Additionally, we measured the complexity of the mechanical signals in terms of statistical significance testing by calculating turning point ratios to detect abnormal episodes in the cardiac signals. Moreover, random behavior in the characteristics of the signal waveform, obtainable by texture spectrum analysis, was assumed to convey information related to the present cardiac mechanical performance. Interested readers are encouraged to see [36] for the mathematical description of the extracted features explained below.

Heart rate (HR) features: HR approximation was achieved by computing short segment auto-correlations from 10 s signal portions. Each 10 s portion was further divided into 2.5 s segments with a 1.5 s overlap (yielded in eight sub-segments). Each sub-segment was cross-correlated within its instant time-domain neighborhood, resulting in eight series of cross-correlation coefficients. From each of the cross-correlation series, the beat time intervals were extracted by computing the deviation of the first side peak. Finally, HR was obtained by calculating the median of these deviations. HRV was obtained by computing the median absolute difference (MEAD) of the obtained beat-to-beat intervals. Furthermore, two additional variants of HRV were formed by calculating the second-order difference and absolute second-order difference of the beat-to-beat intervals [17].

Approximate entropy: Approximate entropy can be used to analyze time-series signal complexity. We used approximate entropy to evaluate the irregularity of the signal segments. First, the signal was down-sampled to 8 Hz to reduce the computation time, and all-against-all matching was performed for short signal sub-segments of length 2. A counter C was updated one-by-one logarithmically ($log(C)$ added at each iteration) based on the number of matching sub-segment pairs C above a threshold. Afterwards, the length of the sub-segment was incremented by one and the procedure repeated. The approximate entropy was calculated by subtracting the normalized second counter result from the first one [37], [17].

Spectral entropy: Spectral entropy measures the randomness of a time series. From the power spectral density, the frequency band in the range from 1 to 11 Hz was extracted. An approximate noise floor removal was then applied by removing frequency components that have an amplitude smaller than $1/6$ of the maximum spectral amplitude. The resulting spectra was normalized to have an area of 1, representing a probability density of P. Finally the spectral entropy was computed with $SP_{ent} = -\sum P(f)log(P(f))$. A larger SP_{ent} value indicate more aperiodic signal, typically for AFib condition [17].

Turning-point ratios: The number of increasing or decreasing signal sub-segments denoted as α were used to compute a turning-point ratio as $TPR(x) = \alpha/(N-2)$, where n was the number of samples in the signal segment. The TPRs were computed with different filtering schemes applied to the signal before the TPR operator, including bandpass filtered versions of the signal with five different passband ranges and the same bandpass filters accompanied by convolution using a long triangular shaped window (to smoothen the signal). In addition, one TPR feature was calculated without filtering with only pre-processing being applied and another was computed based on the beat-to-beat time intervals (total of eight) [17]. As a result, a total of 12 TPR features were found.

Energy features: For each signal segment, 11 energy

features were computed taking the mean square of the signal. The distinct features were obtained by using different bandpass filters with passbands similar to the ones used to derive the TPR features. The energy features were targeted to measure the signal amplitude volume possibly related to either normal or abnormal heart operation.

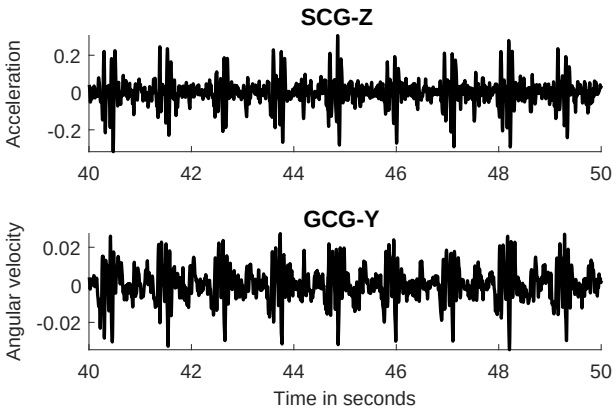
Local binary patterns: Local binary pattern (LBP) features are typically used for texture classification in images. In our case, a 1D variant of LBP was implemented and used [38]. A 9-point kernel centered at a signal value was exploited to form a local neighborhood. Within each local neighborhood, the central point is compared to its neighboring points. For each neighbor, a value is assigned, either 1 if the neighbor is larger or 0 otherwise. Eight bits were generated for each local neighborhood resulting in an LBP. The LBPs were then converted to a decimal. As a final step, a histogram of these decimals in a pre-determined window of the signal (e.g. 10 s) was used as the feature vector. The 8 bit LBP, when converted to decimals, contains 256 bins. We used a subset of these bins, called uniform LBPs [38] by reducing the number of bins to 59. We also applied different distances between the elements/bits of the LBP i.e. the neighboring bits are not directly adjacent but separated by a fixed number of samples. This results in a widening of the window and subsequently a longer part of the signal is covered by a single LBP. In our case, the final length of the LBP histogram was 59, which contains one LBP histogram with a distance of 3 samples.

The overall length of the feature vector is 88 for a single 10 s segment and a single channel. This means that for all the SCG and GCG channels there are altogether $6 * 88(528)$ features which are concatenated to form the final feature vector.

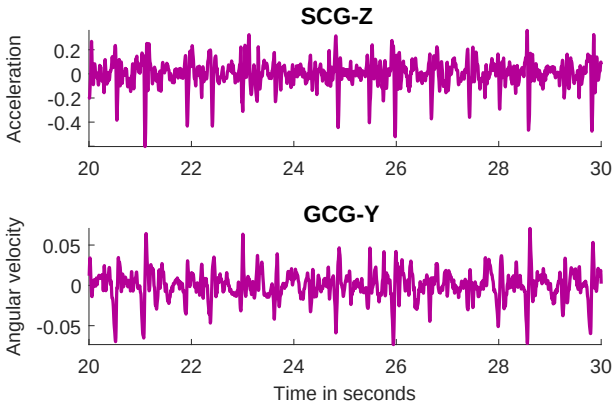
III. Results

Samples of smartphone SCG and GCG cardiac waveforms are plotted in Fig. 4. In this figure, the three different waveforms correspond to the three different disease groups in the analyzed data set. The three categories are (a) controls [non-AFib, non-ADHF], (b) AFib cases who did not have ADHF [AFib, non-ADHF], as well as (c) cases with both AFib and ADHF [AFib, ADHF]. As shown, in the control case, both rotational and translational signals follow a regular rhythm and fairly monomorphic repeating patterns; whilst in [AFib, non-ADHF] and [AFib, ADHF] conditions, cardiac signals appear irregular in terms of rhythm with abnormal morphological characteristics.

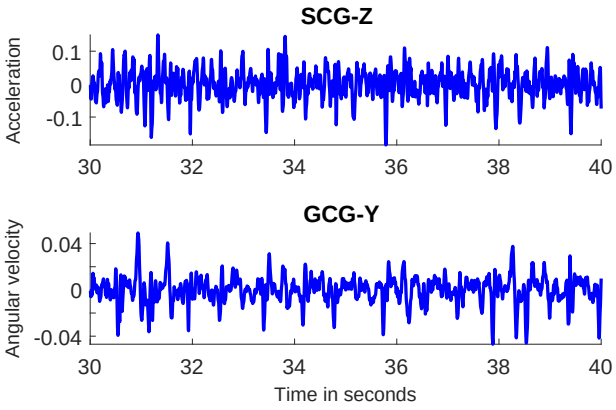
The validation of the classifiers for each disease was performed with nested cross-validation [32]. The receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) were obtained from the nested cross-validation predictions using the pooling approach [39], [40]



(a) True label set: [Non-AFib, non-ADHF].



(b) True label set: [AFib, non-ADHF].

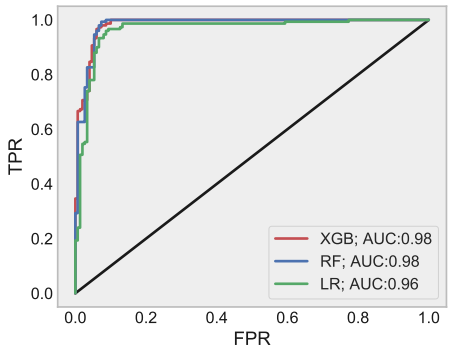


(c) True label set: [AFib, ADHF].

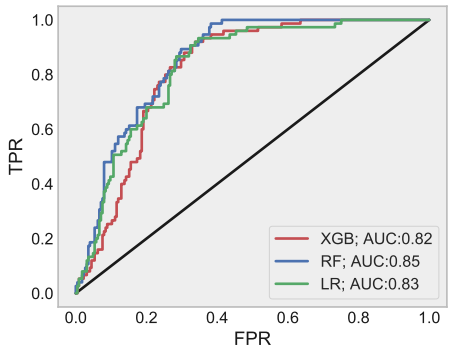
Fig. 4: Example signals of the subjects in the three possible categories namely (a) non-AFib and non-ADHF, (b) AFib and non-ADHF, as well as (c) AFib and ADHF.

A. Multi-label classification approach

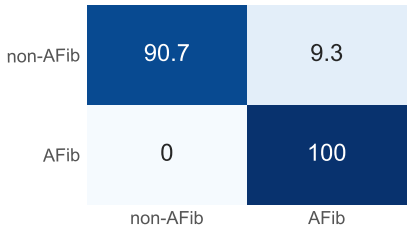
In the multi-label classification approach, the two diseases are classified independently by two two-class classifiers. For each disease, three different classifiers namely RF, XGB, and LR were tested. The ROC curves obtained from the nested cross-validations are shown in Fig. 5a and Fig. 5b. For each disease, there are three ROC curves each corresponding to one of the three classifiers. Figure 5a corresponds to AFib classification and Fig. 5b corresponds to ADHF classification. AUC values of 0.98, 0.98, and 0.96 were achieved by RF, XGB, and LR for AFib; while, the same classifiers resulted in AUC values of 0.85, 0.82, and 0.83 for ADHF, respectively.



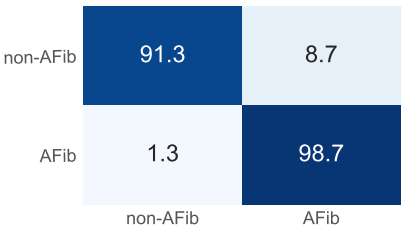
(a) ROC curve for AFib classification.



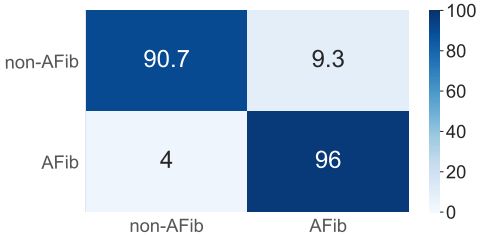
(b) ROC curve for ADHF classification.



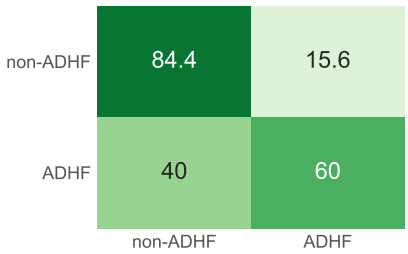
(c) RF for AFib.



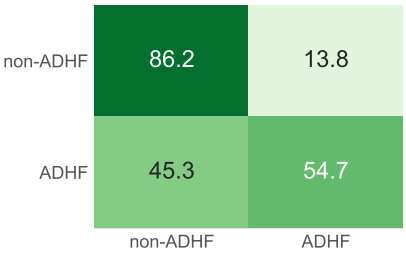
(d) XGB for AFib.



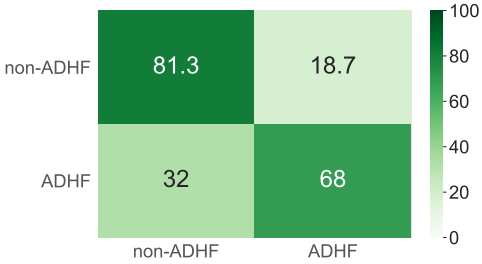
(e) LR for AFib.



(f) RF for ADHF.



(g) XGB for ADHF.



(h) LR for ADHF.

Fig. 5: ROC curves of the independent AFib (a) and ADHF (b) classification in the multi-label classification mode. Normalized confusion matrices (percentages) of the classification performed by three classifiers for AFib (c to e) and ADHF (f to h) are provided.

The confusion matrices for both diseases and the three classifiers can be seen in Fig. 5c to Fig. 5h. Table II summarizes the performance metrics for the two diseases and the deployed classifiers in the multi-label classification approach. The highest percentages of the sensitivity and the specificity values [41] for AFib were 100 and 91.3 obtained by RF and XGB, respectively. The highest percentages of sensitivity and specificity values for ADHF were 68.0 and 86.2 obtained by LR and XGB, respectively. As far as sensitivity is concerned, RF provided the best performance for AFib, whilst LR had the best performance for ADHF.

According to Table II, the highest percentages of positive predictive and negative predictive values for AFib classification were equal to 91.9 and 100 which were obtained by XGB and RF, respectively. The highest percentages of the positive predictive and the negative predictive values for ADHF classification were equal to 56.9 and 88.4 which were obtained by XGB and LR, respectively.

The highest exact match score (EMS), which is defined as the proportion of subjects whose predicted set of labels (both AFib and ADHF) match exactly their set of true labels, was equal to 75.7% using the combination of RF classifier for AFib and either of XGB or LR classifiers for ADHF. The highest micro-average F1-score (F1-micro) [21] was equal to 86.8% obtained by the combination of RF classifier for AFib and either of RF or XGB classifiers for ADHF. The highest macro-average F1-score (F1-macro) was equal to 78.1% obtained by the combination of RF for AFib and LR for ADHF.

B. Hierarchical classification approach

In the case of hierarchical classification, the classification of ADHF was made only for those patients who were predicted as having AFib. The motivation for adopting such an approach is that AFib patients are at a higher risk of acquiring ADHF. Therefore, after a person is classified as AFib, it is crucial to determine whether the patient is also suffering from ADHF. For this purpose, we chose the classifier which provided the best performance on the AFib classification and let that

Disease groups	Classifier	AUC	SE	SP	PPV	NPV
[AFib] vs [non-AFib]	RF	0.98	100	90.7	91.5	100
	XGB	0.98	98.7	91.3	91.9	98.6
	LR	0.96	96	90.7	91.2	95.8
[ADHF] vs [non-ADHF]	RF	0.85	60.0	84.4	56.2	86.4
	XGB	0.82	54.7	86.2	56.9	85.1
	LR	0.83	68.0	81.3	54.8	88.4

TABLE II: Sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) for the two diseases, AFib and ADHF, accomplished by the three deployed classifiers in the multi-label classification framework. Except for the AUC values, all the other values are normalized to percentages.

classify the study subjects into AFib and non-AFib categories. Then, we selected those subjects who fell into the AFib category and proceeded to classify them into ADHF and non-ADHF classes. There were 164 (150 true positives + 14 false positives) AFib cases in our data set who were classified as AFib by the RF classifier (see Fig. 5c). These subjects were, therefore, considered as the target group for ADHF classification. The results obtained from the three deployed classifiers are presented in Figure 6 and Table III. In the case of the hierarchical classification of ADHF, AUC values of 0.64, 0.63, and 0.67 were achieved by RF, XGB, and LR classifiers, respectively.

The highest F1-micro and F1-macro in the hierarchical classification of ADHF was equal to 84.7% and 84.5%, respectively. It is worth noting that in hierarchical classification mode, the EMS is undefined since we only consider a subset of the data.

IV. Discussion

This paper presents two machine learning frameworks to tackle the clinically important issue of simultaneous AFib and ADHF, which is a worldwide cause of morbidity and mortality. We studied the feasibility of using sMCG for supervised multi-label classification as well as supervised hierarchical classification in a patient group that included both single and multi-disease patients. This study was accomplished using only a smartphone without any extra hardware. The built-in inertial sensors, i.e. triaxial accelerometer and gyroscope, of the smartphone were used to measure the precordial translational and rotational micro-vibrations generated from the heart movements in a study sample of 300 cardiac patients. A beneficial application of the multi-label classification of cardiac diseases would be, for example, in the follow-up of high-risk patients, where this method could provide a risk estimate and encourage patients to seek further help and care if needed.

The presented method differs from earlier contributions in that the multi-label classification allows the presence of simultaneous cardiac disorders to be estimated. The more typical multi-class classification approach considers all conditions as mutually exclusive. In the presented case, as illustrated in Fig. 1, 75 of the subjects had two cardiac diseases, highlighting the importance of the multi-label classification approach.

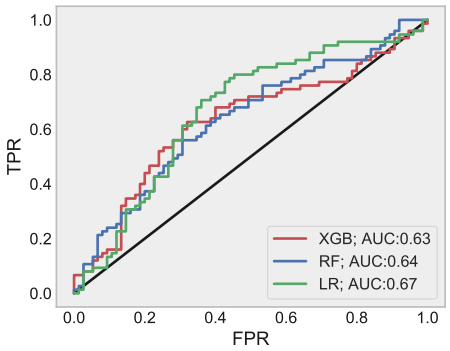
The classifiers were chosen after a thorough experimentation and literature analysis. LR is a simple generalized linear model that has been nonetheless shown to be very competitive in many diagnostic classification tasks [42], [43]. RF [26], [44] and XGB [45] are more advanced state-of-the-art methods that have outperformed other classification methods in many computer science competitions and research studies.

In the multi-label classification scheme, the best classifiers according to accomplished AUC, EMS, and F1-score values are XGB for AFib (Fig. 5d) and LR for ADHF (Fig. 5h), respectively. The best classifier as regards sensitivity and negative predictive value is RF for AFib and LR for ADHF (Table II), respectively. In the hierarchical classification approach, the best classifier was LR according to all the calculated performance metrics. Restricting the ADHF detection to only the AFib patients had no effect on the sensitivity score, however, it resulted in lower levels of specificity, negative predictive value, and AUC. In contrast, the positive predictive value rose to close to 70% in comparison with 56.9% which was achieved by using the multi-label classification approach.

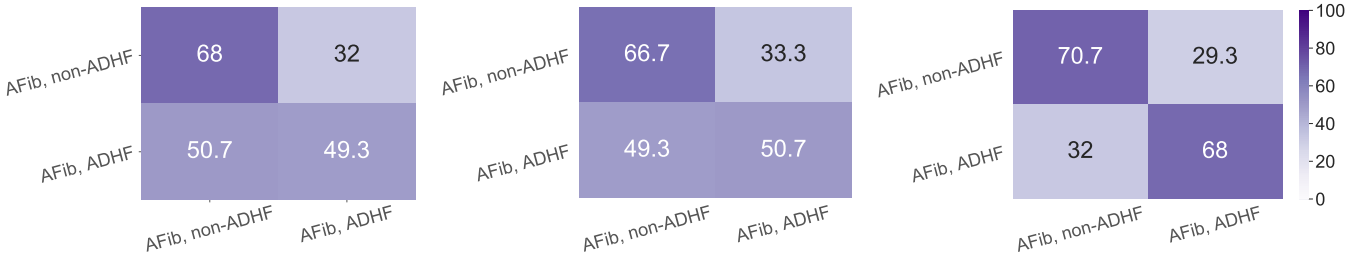
According to the presented results of the multi-label classification framework in Fig. 5, AFib was detected with a comparable sensitivity and specificity values to photoplethysmography-based (PPG) methods, as shown in Table IV. The performance of the proposed AFib classification is also comparable to ECG-based methods as indicated in Table IV. The studies outlined in Table IV were obtained from a literature search among articles published after 2017. An article was selected if it involved a machine learning based classification of AFib and contained performance metrics values. It is worth noting that for a fair comparison of all these studies, sample size variations must be taken into account.

According to the results of multi-label classification framework provided in Fig. 5, ADHF was detected with moderate sensitivity and specificity values but a fairly high negative predictive value. The fairly high negative predictive value for ADHF detection suggests the possibility of using smartphone-derived MCG analysis for ruling out the existence of ADHF. Similar approaches using PPG and ECG signals have been presented by a number of recent studies; these can be accessed in Table V. The most common diagnostic examinations for ADHF are based on echocardiography, BNP and n-terminal-proBNP, ECG, and chest X-ray [5], [60]. Echocardiography is the most feasible and widely used diagnostic tool to demonstrate different cardiac diseases causing the symptoms of clinical ADHF [61]. The other techniques and examinations are mostly used to either complement the findings of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2020.2981334, IEEE Sensors Journal



(a) ROC curve for hierarchical ADHF classification.



(b) RF for ADHF. (c) XGB for ADHF. (d) LR for ADHF.

Fig. 6: ROC curve of the hierarchical classification of ADHF (a) only for AFib category. The corresponding normalized confusion matrices (percentages) of the classification performed by three classifiers for ADHF (b to d) are provided.

Disease groups	Classifier	AUC	SE	SP	PPV	NPV
[AFib, ADHF] vs [AFib, non-ADHF]	RF	0.64	49.3	68.0	60.7	57.3
	XGB	0.63	50.7	66.7	57.5	60.3
	LR	0.67	68	70.7	69.9	68.8

TABLE III: Sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) for hierarchical classification of ADHF, accomplished by the three deployed classifiers. Except for the AUC values, all the other values are normalized to percentages.

echocardiography or provide a diagnosis in the absence of echocardiography [5]. BNP and NT-proBNP showed 100% and 70% sensitivity and specificity values [62]. Therefore, BNP and NT-proBNP are mostly used for ruling out HF, but not establishing a diagnosis [5]. ECG-based methods in contrast provide 77-95% sensitivity and 46-66% specificity values [3], [63], [62]. Examinations based on chest X-ray reported 40-88% sensitivity as well as 66-93% specificity rates [3], [64], [62]. The recently introduced SwellFit, which is a wearable sensor for monitoring congestive HF, has been shown effective in detecting ankle swelling of a few ADHF patients [65]. However, as there has not been any clinical study utilizing that device, the reliability is unknown and remains an open question.

The negative predictive value of ADHF detection considering all 300 patients in this data set, as outlined in the multi-label classification approach, was equal to 88.4%. Whereas, if we only consider the AFib patients – [AFib, ADHF] and [AFib, non-ADHF] categories – the negative predictive value, as outlined with hierarchical classification, decreases to 68.8%. These results imply that in a population of cardiac patients, for those who do not have ADHF, the multi-label classification technique can rule out the presence of ADHF with a fairly high probability. However, if the target population is limited to only AFib patients, ADHF can be ruled out with a lower confidence level mainly due to the greater complexity of the sMCG signals when AFib and ADHF are both present.

AFib detection was performed with high classification accuracy [18]. This is mainly because AFib is an arrhythmia with fully random beat-to-beat interval variations (as is visually observable in Fig. 4b and Fig. 4c). In the case of ADHF, however, no clear morphological patterns – such as beat morphology and/or beat rhythm – can be easily observed by the human eye from the MCG signals.

Considering the number of ADHF cases in this data set and the complexity of ADHF diagnosis, the presented framework reveals a reasonable performance for the ADHF classification. The data set was gathered from elderly adults in a clinical environment. The MCG signals for these individuals are typically weaker than for younger and healthier adults. Therefore, further studies are required to prove the generalizability of the proposed method for the entire population. One future

Author (year) [ref.]	Sample size	Signal type	Methodology	Performance metrics			
				SE	SP	ACC	AUC
Aliamiri et al. (2018) [46]	19	PPG	1D CRNN	-	-	0.982	0.997
Tison et al. (2018) [47]	9801	PPG	1D CRNN	0.980	0.902	-	0.970
Poh et al. (2018) [48]	4386	PPG	1D CNN	0.952	0.990	0.961	0.997
Gotlibovych et al. (2018) [49]	53	PPG	1D CRNN	0.999	0.998	-	0.999
Shashikumar et al. (2018) [50]	2947	PPG	2D CRNN	-	1.00	0.950	0.970
Kwon et al. (2019) [51]	75	PPG	1D CRNN	-	-	0.979	0.998
Fallet et al. (2019) [52]	17	PPG	DT	0.997	0.924	0.981	-
Shen et al. (2019) [53]	81	PPG	1D CNN	-	-	-	0.950
Kora et al. (2017) [54]	44	ECG	LMNN	0.999	0.987	0.993	-
Acharya et al. (2017)[55]	47	ECG	1D CNN	0.991	0.814	0.949	-
Tripathy et al. (2017) [56]	40	ECG	DBN	0.978	0.987	0.983	-
Yao at al. (2017) [57]	608	ECG	1D CNN	0.982	0.981	0.982	0.996
Xia et al. (2018) [58]	>23	ECG	2D CNN	0.988	0.979	0.986	-
Mousavi et al. (2020) [59]	>23	ECG	RNN	0.998	0.985	0.988	0.999

TABLE IV: Recent studies in which PPG-based and ECG-based AFib detection was conducted using machine learning techniques. DBN stands for deep belief networks, DT stands for decision tree, LMNN stands for Levenberg-Marquardt neural network, RNN stands for recurrent neural networks, SVM stands for support vector machines, 1D CNN stands for 1-dimensional CNN, 1D CRNN stands for 1-dimensional CNN coupled with RNN, and 2D CRNN stands for 2-dimensional (image-based) CNN coupled with RNN.

Author (year) [ref.]	Sample size	Signal type	Methodology	Performance metrics			
				SE	SP	ACC	AUC
Baldoumas et al. (2017) [66]	99	PPG	SVM	0.810	0.900	0.930	0.953
Sudarshan et al. (2017) [67]	73	ECG	KNN	0.998	0.999	0.999	-
Acharya et al. (2017) [68]	73	ECG	KNN	0.970	0.982	0.976	-
Al Abdi et al. (2018) [69]	40	ECG	LR	-	-	0.950	-
Bhurane et al. (2019) [70]	76	ECG	SVM	0.998	0.993	0.997	-
Acharya et al. (2019) [71]	73	ECG	1D CNN	0.989	0.990	0.990	-

TABLE V: Recent studies in which PPG-based and ECG-based HF detection was handled using machine learning techniques. KNN denotes K nearest neighbor, LR denotes logistic regression, SVM denotes support vector machines, and 1D CNN stands for 1-dimensional CNN.

direction would be to further evaluate the mechanical motion signals in different clinical settings with even larger study groups. Moreover, no [non-AFib, ADHF] sample was present in this data set therefore this group needs to be considered in future studies. In this study, we used a limited number of smartphone models, and the results, in this case, may only apply to this setting. Further tests are required with other smartphones.

Motion artefact is generally a major issue in MCG signals and needs to be considered in the final application; although, it was not considered a major concern in this study mostly because the data acquisition process was well controlled and supervised. In general, the varying location of the smartphone on the chest may affect, e.g. the amplitude of the acquired signals. In an optimal case, a smartphone application would tolerate major changes in the smartphone placement. On the whole, the most notable limitation of this technique is the necessity of lying in a supine position and remaining motionless during the data collection.

In comparison with other diagnostic examinations, a new recording modality which requires as little professional knowledge and special equipment as possible is favorable as it can be adopted for self-monitoring outside the clinical environment. Self-screening used directly by the patients themselves would be highly beneficial for early interventions.

V. Conclusion

This paper presents a new machine learning paradigm to tackle the clinically important issue of simultaneous AFib and ADHF, which are a major cause of morbidity and mortality worldwide. We employed two state-of-the-art machine learning frameworks, multi-label classification and hierarchical classification, to facilitate the automatic detection of the two cardiac abnormalities in 300 patients. The built-in inertial sensors, i.e. triaxial accelerometer and gyroscope, of the smartphone were used to measure precordial translational and rotational micro-vibrations generated from heart movement. In the paper, we have indicated the feasibility of recognizing simultaneous cardiac abnormalities such as AFib and ADHF using automated analysis of the cardio-mechanical signals.

ACKNOWLEDGEMENTS

The authors would like to thank all the participants of the study. This study was supported partly by the Academy of Finland under grant 290930, the Finnish Foundation for Technology Promotion, as well as the Nokia Foundation.

REFERENCES

[1] E. Anter, M. Jessup, and D. J. Callans, "Atrial fibrillation and heart failure: treatment considerations for a dual epidemic," *Circulation*, vol. 119, no. 18, pp. 2516–2525, 2009.

[2] A. Verma, J. M. Kalman, and D. J. Callans, "Treatment of patients with atrial fibrillation and heart failure with reduced ejection fraction," *Circulation*, vol. 135, no. 16, pp. 1547–1563, 2017.

[3] C. Taylor and R. Hobbs, "Diagnosing heart failure experience and best pathways," *Journal-Echocardiographic Assessment of Diastolic Heart Failure*, 2010.

[4] T. Millane, G. Jackson, C. Gibbs, and G. Lip, "Acute and chronic management strategies," *Bmj*, vol. 320, no. 7234, pp. 559–562, 2000.

[5] N. A.-A. UK, J. J. Atherton, J. Bauersachs, A. J. C. UK, S. Carerj, C. Ceconi, A. Coca, P. E. UK, Ç. Erol, J. Ezekowitz *et al.*, "2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure," *European Heart Journal*, vol. 37, pp. 2129–2200, 2016.

[6] A. Davie, C. Francis, M. Love, L. Caruana, I. Starkey, T. Shaw, G. Sutherland, and J. McMurray, "Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction," *British Medical Journal*, vol. 312, no. 7025, pp. 222–223, 1996.

[7] T. F. Members, G. Montalescot, U. Sechtem, S. Achenbach, F. Andreotti, C. Arden, A. Budaj, R. Bugiardini, F. Crea, T. Cuisset *et al.*, "2013 ESC guidelines on the management of stable coronary artery disease: the task force on the management of stable coronary artery disease of the european society of cardiology," *European Heart Journal*, vol. 34, no. 38, pp. 2949–3003, 2013.

[8] P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. F. Cleland, A. J. S. Coats, V. Falk, J. R. Gonzlez-Juanatey, V.-P. Harjola, E. A. Jankowska, M. Jessup, C. Linde, P. Nihoyannopoulos, J. T. Parissis, B. Pieske, J. P. Riley, G. M. C. Rosano, L. M. Ruilope, F. Ruschitzka, F. H. Rutten, P. van der Meer, and E. S. D. Group, "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," *European Heart Journal*, vol. 37, no. 27, pp. 2129–2200, 05 2016. [Online]. Available: https://doi.org/10.1093/eurheartj/ehw128

[9] M. Etemadi, S. Hersek, J. M. Tseng, N. Rabbani, J. A. Heller, S. Roy, L. Klein, and O. T. Inan, "Tracking clinical status for heart failure patients using ballistocardiography and electrocardiography signal features," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 5188–5191.

[10] B. S. Bozhenko, "Seismocardiography– a new method in the study of functional conditions of the heart," *Terapevticheskii Arkhiv*, vol. 33, pp. 55–64, 1961.

[11] J. Zanetti and D. Salerno, "Seismocardiography: a technique for recording precordial acceleration," in *Computer-Based Medical Systems, 1991. Proceedings of the Fourth Annual IEEE Symposium*, 05 1991, pp. 4–9.

[12] D. M. Salerno, J. M. Zanetti, L. C. Poliac, R. S. Crow, P. J. Hannan, K. Wang, I. F. Goldenberg, and R. A. Van Tassel, "Exercise seismocardiography for detection of coronary artery disease," *American Journal of Noninvasive Cardiology*, vol. 6, pp. 321–330, 1992.

[13] M. Becker, A. Roehl, U. Siekmann, A. Koch, M. de la Fuente, R. Roissant, K. Radermacher, N. Marx, and M. Hein, "Simplified detection of myocardial ischemia by seismocardiography," *Herz*, vol. 39, no. 5, pp. 586–592, 2014.

[14] T. Hurnanen, E. Lehtonen, M. J. Tadi, T. Kuusela, T. Kiviniemi, A. Saraste, T. Vasankari, J. Airaksinen, T. Koivisto, and M. Pänkäälä, "Automated detection of atrial fibrillation based on time–frequency analysis of seismocardiograms," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1233–1241, 2017.

[15] O. T. Inan, M. Baran Pouyan, A. Q. Javaid, S. Dowling, M. Etemadi, A. Dorier, J. A. Heller, A. O. Bicen, S. Roy, T. De Marco *et al.*, "Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients," *Circulation: Heart Failure*, vol. 11, no. 1, p. e004313, 2018.

[16] M. J. Tadi, E. Lehtonen, A. Saraste, J. Tuominen, J. Koskinen, M. Teräs, J. Airaksinen, M. Pänkäälä, and T. Koivisto, "Gyrocardiography: A new non-invasive monitoring method for the assessment of cardiac mechanics and the estimation of hemodynamic variables," *Scientific Reports*, vol. 7, no. 1, p. 6823, 2017.

[17] O. Lahdenoja, T. Hurnanen, Z. Iftikhar, S. Nieminen, T. Knuutila, A. Saraste, T. Kiviniemi, T. Vasankari, J. Airaksinen, M. Pänkäälä *et al.*, "Atrial fibrillation detection via accelerometer and gyroscope of a smartphone," *IEEE Journal of Biomedical and Health Informatics*, 2017.

[18] J. Jaakkola, S. Jaakkola, O. Lahdenoja, T. Hurnanen, T. Koivisto, M. Pänkäälä, T. Knuutila, T. O. Kiviniemi, T. Vasankari, and K. J. Airaksinen, "Mobile phone detection of atrial fibrillation with mechanocardiography: the mode-af study (mobile phone detection of atrial fibrillation)," *Circulation*, vol. 137, no. 14, pp. 1524–1527, 2018.

[19] O. Lahdenoja, T. Koivisto, M. J. Tadi, Z. Iftikhar, T. Hurnanen, T. Vasankari, T. Kiviniemi, J. Airaksinen, and M. Pänkäälä, "A smartphone-only solution for detecting indications of acute myocardial infarction," in *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*. IEEE, 2017, pp. 197–200.

[20] S. Mehrang, M. J. Tadi, M. Kaisti, O. Lahdenoja, T. Vasankari, T. Kiviniemi, J. Airaksinen, T. Koivisto, and M. Pänkäälä, "Machine learning based classification of myocardial infarction conditions using smartphone-derived seismo-and gyrocardiography," in *2018 Computing in Cardiology Conference (CinC)*, vol. 45. IEEE, 2018, pp. 1–4.

[21] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine Learning and Knowledge Discovery in Databases*, pp. 254–269, 2009.

[22] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, 2006.

[23] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014.

[24] O. Luaces, J. Díez, J. Barranquero, J. J. del Coz, and A. Bahamonde, "Binary relevance efficacy for multilabel classification," *Progress in Artificial Intelligence*, vol. 1, no. 4, pp. 303–313, 2012.

[25] C. N. Silla and A. A. Freitas, "A survey of hierarchical classification across different application domains," *Data Mining and Knowledge Discovery*, vol. 22, no. 1-2, pp. 31–72, 2011.

[26] M. J. Tadi, S. Mehrang, M. Kaisti, O. Lahdenoja, T. Hurnanen, J. Jaakkola, S. Jaakkola, T. Vasankari, T. Kiviniemi, J. Airaksinen *et al.*, "Comprehensive analysis of cardiogenic vibrations for automated detection of atrial fibrillation using smartphone mechanocardiograms," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2230–2242, 2018.

[27] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2017, pp. 141–144.

[28] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[29] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[30] S. J. Press and S. Wilson, "Choosing between logistic regression and discriminant analysis," *Journal of the American Statistical Association*, vol. 73, no. 364, pp. 699–705, 1978.

[31] M. Esterman, B. J. Tamber-Rosenau, Y.-C. Chiu, and S. Yantis, "Avoiding non-independence in fMRI data analysis: leave one subject out," *Neuroimage*, vol. 50, no. 2, pp. 572–576, 2010.

[32] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, p. 91, 2006.

[33] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[35] A. Lay-Ekuakille, P. Vergallo, A. Trabacca, M. De Rinaldis, F. Angelillo, F. Conversano, and S. Casciaro, "Low-frequency detection in ECG signals and joint eeg-ergospirometric measurements for precautionary diagnosis," *Measurement*, vol. 46, no. 1, pp. 97–107, 2013.

[36] Z. Iftikhar, O. Lahdenoja, M. J. Tadi, T. Hurnanen, T. Vasankari, T. Kiviniemi, J. Airaksinen, T. Koivisto, and M. Pänkäälä, "Multiclass classifier based cardiovascular condition detection using smartphone mechanocardiography," *Scientific Reports*, vol. 8, no. 1, p. 9344, 2018.

[37] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[38] N. Chatlani and J. J. Soraghan, "Local binary patterns for 1-D signal processing," in *Signal Processing Conference, 2010 18th European*. IEEE, 2010, pp. 95–99.

[39] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[40] A. Airola, T. Pahikkala, W. Waegeman, B. De Baets, and T. Salakoski, "An experimental comparison of cross-validation techniques for estimating the area under the roc curve," *Computational Statistics & Data Analysis*, vol. 55, no. 4, pp. 1828–1844, 2011.

[41] A. G. Lalkhen and A. McCluskey, "Clinical tests: sensitivity and specificity," *Continuing Education in Anaesthesia Critical Care & Pain*, vol. 8, no. 6, pp. 221–223, 2008.

[42] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 113–127, 2005.

[43] I. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," *Expert Systems with Applications*, vol. 34, no. 1, pp. 366–374, 2008.

[44] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[45] S. D. Goodfellow, A. Goodwin, R. Greer, P. C. Laussen, M. Mazwi, and D. Eytan, "Classification of atrial fibrillation using multidisciplinary features and gradient boosting," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[46] A. Aliamiri and Y. Shen, "Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2018, pp. 442–445.

[47] G. H. Tison, J. M. Sanchez, B. Ballinger, A. Singh, J. E. Olgin, M. J. Pletcher, E. Vittinghoff, E. S. Lee, S. M. Fan, R. A. Gladstone *et al.*, "Passive detection of atrial fibrillation using a commercially available smartwatch," *JAMA cardiology*, vol. 3, no. 5, pp. 409–416, 2018.

[48] M.-Z. Poh, Y. C. Poh, P.-H. Chan, C.-K. Wong, L. Pun, W. W.-C. Leung, Y.-F. Wong, M. M.-Y. Wong, D. W.-S. Chu, and C.-W. Siu, "Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms," *Heart*, vol. 104, no. 23, pp. 1921–1928, 2018.

[49] I. Gotlibovych, S. Crawford, D. Goyal, J. Liu, Y. Kerem, D. Benaron, D. Yilmaz, G. Marcus, and Y. Li, "End-to-end deep learning from raw sensor data: Atrial fibrillation detection using wearables," *arXiv preprint arXiv:1807.10707*, 2018.

[50] S. P. Shashikumar, A. J. Shah, G. D. Clifford, and S. Nemati, "Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 715–723.

[51] S. Kwon, J. Hong, E.-K. Choi, E. Lee, D. E. Hostallero, W. J. Kang, B. Lee, E.-R. Jeong, B.-K. Koo, S. Oh *et al.*, "Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: Algorithms development study," *JMIR mHealth and uHealth*, vol. 7, no. 6, p. e12770, 2019.

[52] S. Fallet, M. Lemay, P. Renevey, C. Leupi, E. Pruvot, and J.-M. Vesin, "Can one detect atrial fibrillation using a wrist-type photoplethysmographic device?" *Medical & Biological Engineering & Computing*, vol. 57, no. 2, pp. 477–487, 2019.

[53] Y. Shen, M. Voisin, A. Aliamiri, A. Avati, A. Hannun, and A. Ng, "Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1909–1916.

[54] P. Kora, A. Annavarapu, P. Yadlapalli, K. S. R. Krishna, and V. Somalaraju, "ECG based atrial fibrillation detection using sequency ordered complex hadamard transform and hybrid firefly algorithm," *Engineering Science and Technology, an International Journal*, vol. 20, no. 3, pp. 1084–1091, 2017.

[55] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017.

[56] R. Tripathy, M. R. A. Paternina, J. G. Arrieta, and P. Pattanaik, "Automated detection of atrial fibrillation ECG signals using two stage vmd and atrial fibrillation diagnosis index," *Journal of Mechanics in Medicine and Biology*, vol. 17, no. 07, p. 1740044, 2017.

[57] Z. Yao, Z. Zhu, and Y. Chen, "Atrial fibrillation detection by multi-scale convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–6.

[58] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 93, pp. 84–92, 2018.

[59] S. Mousavi, F. Afghah, and U. R. Acharya, "Han-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks," *arXiv preprint arXiv:2002.05262*, 2020.

[60] F. Hobbs, R. Davis, A. Roalfe, R. Hare, M. Davies, and J. Kenkre, "Reliability of n-terminal pro-brain natriuretic peptide assay in diagnosis of heart failure: cohort study in representative and high risk community populations," *Bmj*, vol. 324, no. 7352, p. 1498, 2002.

[61] J. N. Kirkpatrick, M. A. Vannan, J. Narula, and R. M. Lang, "Echocardiography in heart failure: applications, utility, and new horizons," *Journal of the American College of Cardiology*, vol. 50, no. 5, pp. 381–396, 2007.

[62] J. Mant, J. Doust, A. Roalfe, P. Barton, M. Cowie, P. Glasziou, D. Mant, R. McManus, R. Holder, J. Deeks *et al.*, "Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care," in *NIHR Health Technology Assessment programme: Executive Summaries*. NIHR Journals Library, 2009.

[63] K. Dickstein, A. F. Members, A. Cohen-Solal, G. Filippatos, J. J. McMurray, P. Ponikowski, P. A. Poole-Wilson, A. Strömberg, D. J. van Veldhuisen, D. Atar *et al.*, "ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: The task force for the diagnosis and treatment of acute and chronic heart failure 2008 of the european society of cardiology. developed in collaboration with the heart failure association of the ESC (HFA) and endorsed by the european society of intensive care medicine (esicm)," *European Journal of Heart Failure*, vol. 10, no. 10, pp. 933–989, 2008.

[64] M. R. Zile and D. L. Brutsaert, "New concepts in diastolic dysfunction and diastolic heart failure: Part I: diagnosis, prognosis, and measurements of diastolic function," *Circulation*, vol. 105, no. 11, pp. 1387–1393, 2002.

[65] S. Kim, Y. S. Iravantchi, K. Z. Gajos, and B. J. Grosz, "Swellfit: a wearable sensor for patients with congestive heart failure," in *Proceedings of the Workshop on Interactive Systems in Healthcare*, 2016.

[66] G. Baldoumas, D. Peschos, G. Tatsis, S. K. Chronopoulos, V. Christofilakis, P. Kostarakis, P. Varotsos, N. V. Sarlis, E. S. Skordas, A. Bechlioulis *et al.*, "A prototype photoplethysmography electronic device that distinguishes congestive heart failure from healthy individuals by applying natural time analysis," *Electronics*, vol. 8, no. 11, p. 1288, 2019.

[67] V. K. Sudarshan, U. R. Acharya, S. L. Oh, M. Adam, J. H. Tan, C. K. Chua, K. P. Chua, and R. San Tan, "Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals," *Computers in Biology and Medicine*, vol. 83, pp. 48–58, 2017.

[68] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. L. Oh, A. Muhammad, J. E. Koh, J. H. Tan, C. K. Chua, K. P. Chua, and R. San Tan, "Application of empirical mode decomposition (EMD) for automated identification of congestive heart failure using heart rate signals," *Neural Computing and Applications*, vol. 28, no. 10, pp. 3073–3094, 2017.

[69] R. M. Al Abdi and M. Jarrah, "Cardiac disease classification using total variation denoising and morlet continuous wavelet transformation of ECG signals," in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, 2018, pp. 57–60.

[70] A. A. Bhurane, M. Sharma, R. San-Tan, and U. R. Acharya, "An efficient detection of congestive heart failure using frequency localized filter banks for the diagnosis with ECG signals," *Cognitive Systems Research*, vol. 55, pp. 82–94, 2019.

[71] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. San Tan, "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals," *Applied Intelligence*, vol. 49, no. 1, pp. 16–27, 2019.