IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 4, 2019, accepted September 24, 2019, date of publication October 7, 2019, date of current version October 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946117

Reliability of Self-Applied Smartphone Mechanocardiography for Atrial Fibrillation Detection

SAEED MEHRANG[1], MOJTABA JAFARI TADI[1], TERO HURNANEN[1], TIMO KNUUTILA[1], OLLI LAHDENOJA[1], JUSSI JAAKKOLA[2], SAMULI JAAKKOLA[2], TUIJA VASANKARI[2], TUOMAS KIVINIEMI[2], JUHANI AIRAKSINEN[2], TERO KOIVISTO[1], AND MIKKO PÄNKÄÄLÄ[1]

[1]Department of Future Technologies, University of Turku, 20500 Turku, Finland
[2]Heart Center, Turku University Hospital, 20500 Turku, Finland

Corresponding author: Saeed Mehrang (saeed.mehrang@utu.fi)

This work was supported by the Academy of Finland under Grant 290930.

ABSTRACT Smartphone mechanocardiography (sMCG) is a new technique that allows patients to record a cardiac rhythm strip using a smartphone built-in tri-axial accelerometer and gyroscope. In this study, we investigated how a self-applied sMCG can reliably contribute to the differentiation of atrial fibrillation (AFib) from the sinus rhythm (SR). A study sample of 300 elderly adults including 150 AFib cases with persistent and paroxysmal AFib was recruited. Among the 300 subjects, 182 subjects (82 AFib) completed two recordings, one physician-applied and one self-applied. The remaining patients (n = 118) were nervous, in a quite poor condition, and not interested in or capable of concentrating on using a smartphone in the acute setting of a hospital in order to perform the self-applied recording. Two data processing frameworks were used, knowledge-based learning (KL) which is a rule-based algorithm and machine learning (ML) which is an automated classification technique. For the ML approach, we considered four classifiers, namely random forest (RF), extreme gradient boosting (XGB), support vector machines (SVM), and artificial neural network (NN). For evaluation, a leave-one-subject-out cross-validation was adopted for the ML approach. Compared to physician-interpreted ECG-derived labels, the KL approach predicted AFib with sensitivity values of 0.963 and 0.976, specificity values of 0.980 and 0.929, and F-measure values of 0.972 and 0.952 for the physician- and self-applied measurements, respectively. Similarly, NN which was the best ML classifier according to the F-measure values, demonstrated, on average, sensitivity values of 0.976 and 0.938, specificity values of 0.962 and 0.936, and F-measure values of 0.969 and 0.937, respectively. All other classifiers delivered quite similar results. The sMCG technology for AFib detection, supported by the KL and ML approaches, can accurately differentiate AFib from SR in both physician- and self-applied recording scenarios. This new technology can help to screen patients with episodic or undiagnosed AFib and also be used as a home-based self-applied monitoring technique.

INDEX TERMS Accelerometers, biomedical engineering, cardiography, gyroscopes, machine learning, signal processing.

I. INTRODUCTION

Atrial fibrillation (AFib) is one of the most prevalent cardiovascular conditions that can result in a stroke and heart failure, two of the most common causes of mortality and morbidity [1]. An irregular heart rhythm as well as fast atrial pacing, not always detectable with suboptimal heart monitoring modalities, are the major characteristics of AFib. Diagnosis of AFib is complicated since the shape and the chronology of the next ventricular complex cannot be predicted. Detecting inter-beat intervals that are completely random and irregular utilizing a rhythm irregularity analysis is a favorable approach to AFib diagnosis [2]. Abnormal ventricular complexes (due to the rapid atrial pacing more

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi.

than 400 bpm) are recognizable in episodes where amplitude and morphology of the intraventricular conduction and fusion alternate with a fast or slow frequency. Detecting irregular variability of the ventricular complexes is a well-grounded approach to a reliable AFib diagnosis [2].

Diagnosing asymptomatic and/or paroxysmal AFib is known to be very challenging as episodes of AFib start occasionally and stop spontaneously [3]. These infrequently appearing AFib episodes may or may not be seen in a short electrocardiogram (ECG) strip taken during a routine checkup at a clinic. In the early stages of developing AFib, it may remain undetected since the patient might not feel any symptoms; this is called silent AFib [1]. Home-monitoring via ubiquitous and/or hand-held recording devices could help to recognize paroxysmal AFib by gathering more frequent and longer (intermittent) snapshots of beat-to-beat intervals during day-to-day life [2]. Nevertheless, self-recording of vital signs and symptoms, and self-interpretation of the acquired data remain open challenges that need to be addressed.

Today, self-monitoring of cardiovascular health status has become increasingly straightforward by the advent of smartphones, smartwatches, and wearable sensors. Sustainable ambulatory AFib monitoring modalities require no more devices than a single smartphone, smartwatch, or a comfortable wearable/handheld sensor. Clinically validated devices such as smartphone- and smartwatch-coupled ECG monitors (e.g. AliveCore Kardiamobile, Kardiaband [4], Apple Watch [5], [6]), hand-held ECG recorders (e.g. Zenicor EKG [7] and Mydiagnostick [8]), the wearable ECG patch (Zio Patch [9]), wearable/mobile photoplethysmography (PPG) recorders (e.g. Fibricheck [10], Preventicus [11], and Samsung Simband [12]) can successfully enable self-detection of heart arrhythmia and in particular sporadic AFib outside clinical settings. However, AFib may remain sporadic and unrecognizable even with the above-mentioned devices. This implies that long-term continuous monitoring techniques are required for an in-time diagnosis [13]. To detect symptoms appearing at periodic or random intervals, an effective strategy for longer-term monitoring, up to several days or weeks at a time, is needed [13], [14]. Despite vigorous research, efficient and cost-effective approaches to detect and screen for asymptomatic AFib are yet to be introduced.

Smartphone mechanocardiography (sMCG) is a new heart monitoring technique which records precordial chest acceleration (seismocardiogram or SCG) and angular velocity (gyrocardiogram or GCG) signals. This new technique allows patients to record a cardiac rhythm strip using smartphone built-in inertial measurement units (IMU) including tri-axial microelectromechanical (MEMS) accelerometer and gyroscope sensors. In this study, we sought to investigate the feasibility of self-monitoring using sMCG for AFib detection for clinical patients. In our previous contributions [2], [15], [16], we introduced mobile phone detection of AFib by processing only physician-applied sMCG recordings using different data analytic architectures. *In this work, however, we aim to further evaluate diagnostic accuracy and reliability of the AFib detection using a self-applied sMCG recording*. To this end, we obtained sMCG data from a group of subjects (n = 300) who were examined during two recording scenarios. Among the 300 subjects, 182 subjects (82 AFib) completed two recordings, one physician-applied and one self-applied sMCG recording. The remaining subjects (n = 118) were either unable or unwilling to perform the self-applied recording. It must be taken into account that our smartphone app was not designed for elderly users at the time of recording. Besides, many elderly users had no previous experience of using smartphone ehealth apps. However, not a single recording was rejected because of poor signal quality.

Diagnostic reliability of sMCG for AFib detection was assessed by deploying two independent approaches, namely knowledge-based learning (KL) and machine learning (ML). The KL algorithm, which operates based on characterizing the regularity of the beating pattern of the heart, was deployed to predict AFib in those subjects who had completed both physician- and self-applied measurements (n = 182). For the ML approach, training and testing the ML algorithm was done through an iterative process in which every time one subject's measurements were placed into the test set and the rest of the measurements into the training set. The testing process was only done for the same aforementioned 182 subjects. The ML approach was examined using four different classifiers, namely random forest (RF) [17], extreme gradient boosting (XGB) [18], support vector machines (SVM) [19], and artificial neural network (NN) [20]. Then the intra- and inter-agreement of the predictions of the two algorithms were assessed for each subject and each category. To facilitate the assessment of subject-by-subject predictions, we introduce barcode plots with colored ribbons to mark the true label and the predicted labels of the measurements of each subject.

II. MATERIALS AND METHODS

A. STUDY PARTICIPANTS

The study sample considered in this paper includes clinical patients from the MODE-AF (mobile-phone detection of atrial fibrillation) trial. The MODE-AF study (ClinicalTrials.gov Identifier: NCT03274583) examined the feasibility and effectiveness of AFib detection with a smartphone and a complementary app for acquiring MCG signals at a 200 Hz sampling frequency. The smartphone was placed on the bare chest of the participants without any chest belt, strap, or any additional sensor attachment equipment. The smartphone was oriented with its screen facing upward and its lower edge aligned with the lower edge of the sternum. For the clinical trial, 300 (150 persistent and paroxysmal AFib) participants with a previous history of cardiovascular diseases with a mean age of 74.75 (range 73.0 to 76.6) were enrolled and their written informed consent obtained. Detail descriptions of the measurement protocol and the demographics of the participants are available in [2], [16].

A trained physician was responsible for the recruitment of all the subjects, obtaining all the sMCG recordings, and performing physician-applied measurements and data documentation. Of the 300 subjects, 193 subjects were also able to collect at least one self-applied recording successfully. However, 11 subjects who collected less than 30 seconds self-applied data were excluded.

The study administrator instructed the subjects on how to turn the device on, how to navigate to the data acquisition application and how to start the recording. They were then instructed to find the correct placement of the phone on their chest. The instruction was repeated as many times as needed individually. After the investigator deemed that the subject had understood the task sufficiently, the first recording was performed. Each subject was given three chances to collect at least one successful self-applied measurement. For every attempt, the subjects were left alone for 10 minutes to carry out the recording on their own. The longest recording was then selected for the analysis. The investigator evaluated subjectively how well the subjects had learned to use the application or whether they had refused to attempt the recording.

Altogether, 182 (82 AFib) subjects whose both physician- and self-applied recordings were longer than 30 seconds were considered suitable for the study. The remaining subjects who were either unable or unwilling to carry out the self-applied recording according to the given instructions were excluded from the prospective head-to-head analysis. Corresponding physician-applied recordings of these excluded subjects were however used for training the implemented ML models to classify AFib, as described in the following sections.

III. ANALYSIS PIPELINES

For this study, we evaluated the detection performance of two independent classification techniques, namely machine learning (or ML) and knowledge-based learning (or KL). The KL approach thoroughly examines the presence of periodicity in the recorded sMCG signals, i.e. SCG and GCG, while the ML approach builds a model utilizing various time-frequency features extracted from these signals.

A. KL APPROACH

The KL approach is a rule-based algorithm which operates based on a sequence of rather simple but effective tests, which are used to determine whether the signal is periodic (i.e. SR) or not (i.e. AFib). A flowchart of the procedure exploited in the KL approach is presented in Figure 1. The ventricular event rate of a normal heart is typically lower than the high atrial event rate of the heart in AFib. Hence, screening of AFib is possible simply by monitoring the overall temporal periodicity of MCG signals. To this end, signals are preprocessed for baseline wander and noise removal by bandpass filtering (passband 4-40 Hz). In spite of the high lower passband limit, the periodicity information of the signal is still retained due to the impulsive nature of the heart signal.

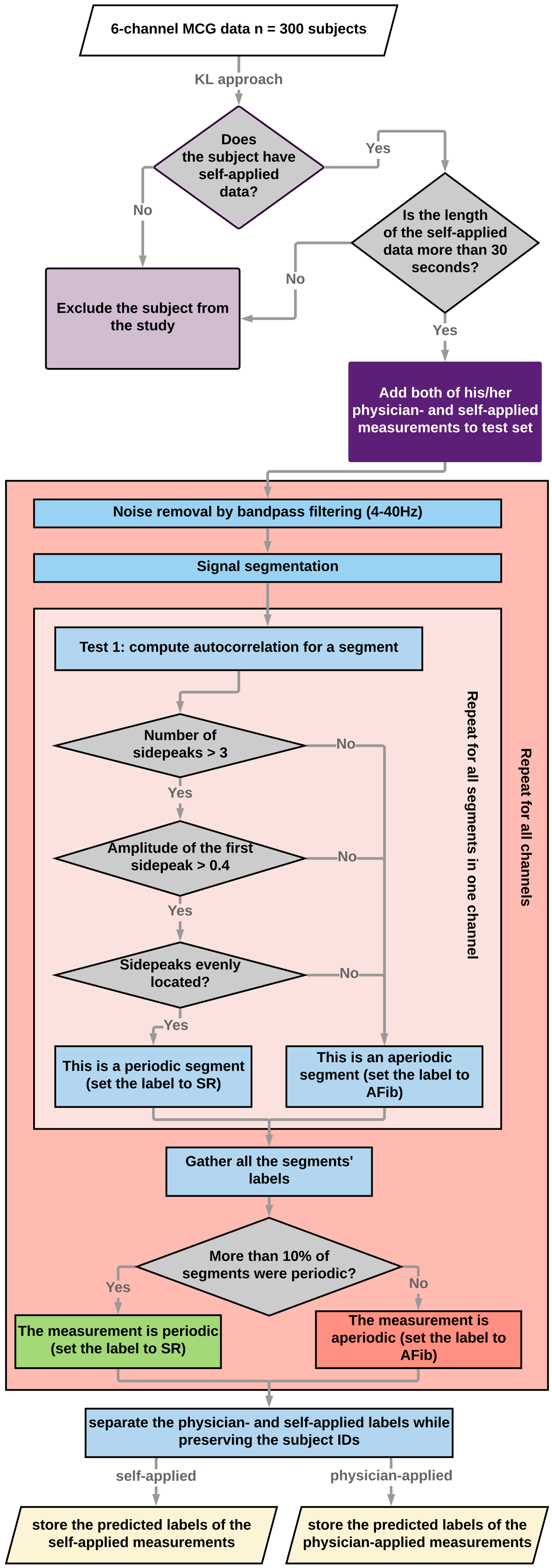


FIGURE 1. The flowchart of the analysis pipeline designed for the KL approach.

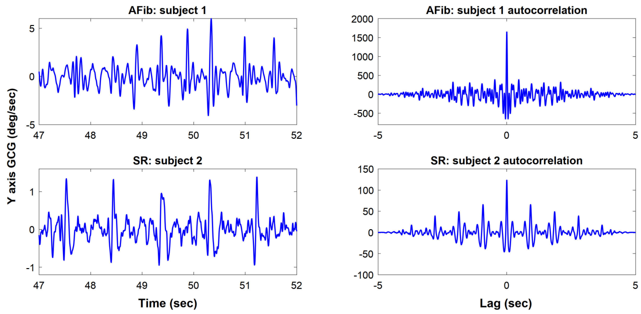


FIGURE 2. The autocorrelation of GCG-Y of a typical AFib measurement (top row) and a typical SR measurement (bottom row). The measurements correspond to the same subjects whose data is shown in Figure 6 top two rows.

The pre-processed signals in all channels were then segmented temporally into 7-second segments with a 2-second overlap, within which periodicity information was obtained for each channel separately. The segment length and the overlap size were selected empirically.

The periodicity of every signal segment was obtained using autocorrelation as described in the following. For each 7-second segment, autocorrelation peaks were found. Based on the characteristics of the peaks three successive tests were implemented to exclude aperiodic signals. The first test determines whether the number of autocorrelation side-peaks is at least three in the current channel and the current 7-second segment. If this is true, the channel in question is a candidate for being periodic (otherwise aperiodic). Second, the amplitude of the first side-peak needs to be at least 0.4, to retain the channel as a candidate for being periodic (otherwise it is aperiodic). In the third phase, we tested whether all the side-peaks were evenly located. If all these requirements are fulfilled for at least one channel, the whole 7-second segment is labelled as periodic. The state of periodicity is then denoted by $0s$ for aperiodic and $1s$ for periodic segments.

After going through all the segments, if the number of periodic segments is above 10% of the total number of segments, the measurement is considered periodic (i.e. SR) and aperiodic otherwise (i.e. AFib). Figure 2 illustrates the application of autocorrelation for quantification of beat-to-beat regularity. The top row in this figure corresponds to an AFib case while the bottom row corresponds to an SR case. The autocorrelation of the AFib case lacks consistent side-peaks while for the SR case there is a clear repeating pattern.

Both the physician- and self-applied measurements in the study test set were evaluated by the KL approach and thereafter their predicted labels were obtained accordingly.

B. ML APPROACH

The ML approach consists of three major building blocks namely signal pre-processing, feature extraction, and classification. A visual step-by-step procedure implemented in the ML approach is presented in Figure 3.

1) SIGNAL PRE-PROCESSING

Singular spectrum analysis (SSA) [21] was exploited to discard noisy components of sMCG signals. In practice, SSA generates a trajectory matrix from the original series of sMCG signals by a sliding window of length $\approx$ 100ms. The trajectory matrix is approximated using singular value decomposition. Next, the components of the trajectory matrix that were less affected by the motion artefacts and noise are found by visual inspection and contribute to reconstructing the noise-free signals. The SSA function applies signal smoothing, filtering, and detrending as described in [16]. Following the SSA process, the envelopes of the signals were extracted to be used for feature extraction. The envelops were computed by adopting three-stage, moving-average filtering with a growing window size [16]. As a result of this pre-processing, two pre-processed channels were developed for every channel in the raw data; that is, in total six SSA-filtered and six envelopes of SSA filtered channels were formed and subsequently fed into the feature extraction phase. Before the feature extraction, signals were segmented into 10-second segments with a 75% overlap. Overlapping segments provide a smoother change of the parameters as the frames move forward.

2) FEATURE EXTRACTION

A wide spectrum of time-frequency features was extracted from both SSA-filtered and corresponding MCG signal envelopes. Table 1 displays the types of features which were computed from each segment of every SCG and GCG channel. Since we were aiming to predict a single label for each measurement rather than each signal segment, we took the median of the calculated features over all the signal segments. This allowed us to obtain a robust overview of the computed features in each measurement [12], [16]. In total, 552 features were extracted for each measurement. The detailed explanation of the mathematics behind each feature is described in [16].

3) TRAINING MODEL AND VALIDATION

Since we consider a supervised learning approach for AFib detection, the learner is provided with two sets of data, a training set and a test set. Training and testing the ML algorithm were done through an iterative process in which every time one subject's measurements were placed into the test set and the other subjects' measurements were placed into the training set (see Figure 3). The testing process was only done for those subjects (n = 182) who had both physician- and self-applied measurements. Those subjects who only had physician-applied measurements were always present in the training set and never present in the test set. For instance, in iteration one, one subject was randomly selected from those who had both physician- and self-applied measurements and was regarded as a test subject. The training set was subsequently comprised of the remaining 299 physician-applied and the remaining 181 self-applied measurements.

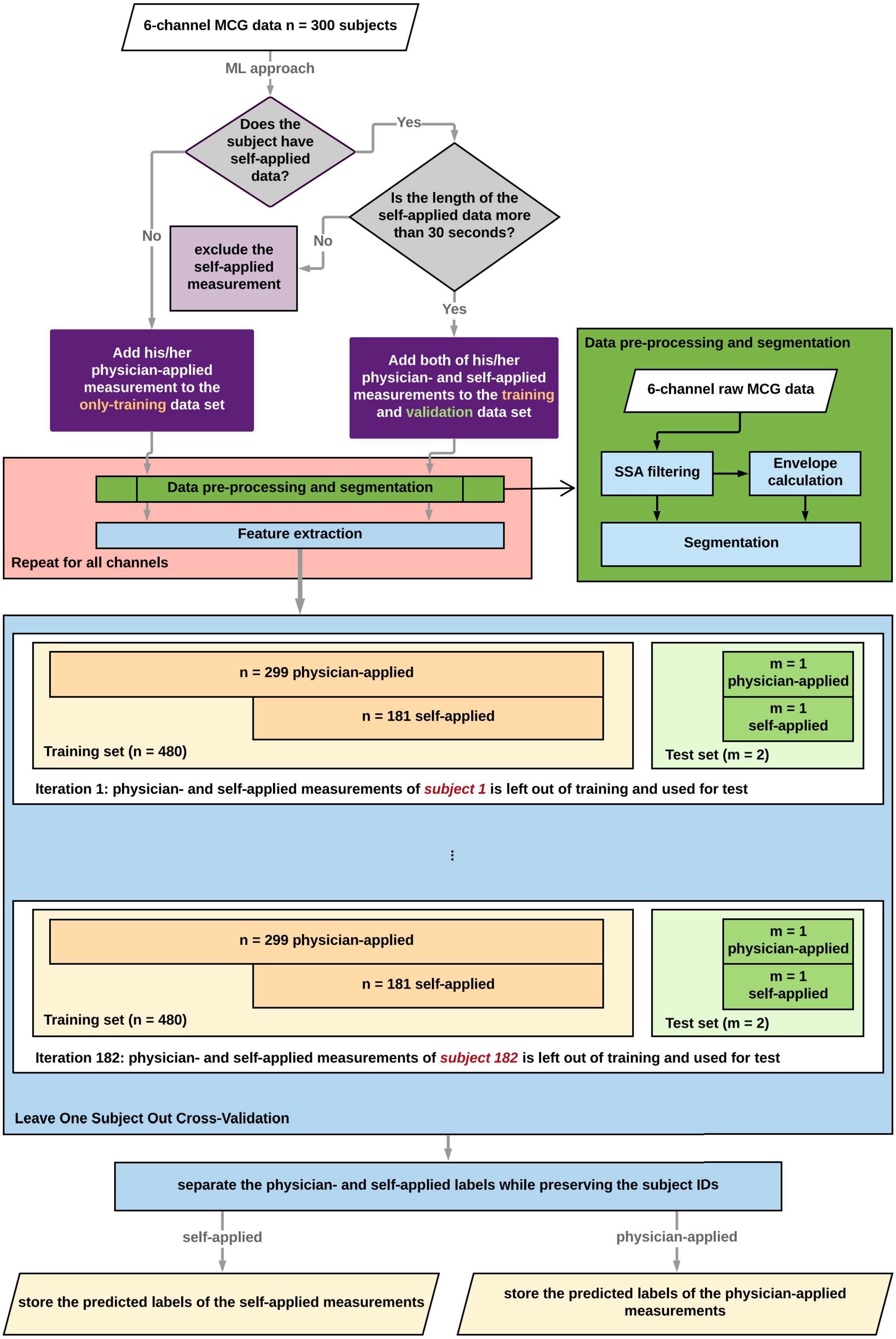


FIGURE 3. The flowchart of the analysis pipeline designed for the ML approach.

TABLE 1. The extracted features were adopted from the Jafari et al. study [16].

Number	Feature	Number	Feature
1	Zero crossing rate (ZCR)	29-31	Covariance
2	Signal energy (E_s)	32	Approximate entropy (ApEn)
3	Signal energy entropy (Ent_e)	33	Spectral entropy (SpEn)
4	Spectral centroid ($SP_{centroid}$)	34	Shannon entropy (SEE)
5	Spectral spread (SP_{spread})	35	Turning point ratios (TPR)
6	Spectral flux (SP_{flux})	36	Median of RR intervals
7	Spectral roll-off	37	RMSSD
8	Fundamental frequency (F_0)	38	PSD of RR intervals
9	Heart rate (HR)	39	Higuchi fractal dimensions (HFD)
10	Harmonic-to-noise ratio (HNR)	40-41	Hjorth parameters
11-16	6 dominant spectral peaks	42	Sample entropy
17-22	Corresponding frequencies of the 6 spectral peaks	43-44	Kurtosis and skewness
23-27	Power spectral density (PSD) in 5 frequency bands	45-46	Arithmetic mean and RMS level
28	Median amplitude spectrum (MAD)		

TABLE 2. Prediction performances of the KL and the ML approaches for the physician- and self-applied data. The presented numbers correspond to the comparison of true labels and the predicted labels.

Measurement Status	Approach	Classifier	Sensitivity	Specificity	F-measure	Kappa-score	AUC
Self-applied	ML	RF	**0.948 (0.01)**	0.920 (<0.01)	0.934 (<0.01)	0.864 (0.01)	0.975 (<0.01)
		XGB	0.915 (0.00)	0.920 (<0.01)	0.917 (<0.01)	0.834 (0.00)	**0.979 (0.00)**
		SVM	0.927 (<0.01)	0.920 (<0.01)	0.923 (<0.01)	0.845 (0.00)	0.978 (<0.01)
		NN	0.938 (0.01)	**0.936 (0.01)**	**0.937 (0.01)**	**0.873 (0.01)**	0.973 (0.01)
	KL		**0.976 (0.00)**	**0.929 (0.00)**	**0.952 (0.00)**	**0.900 (0.00)**	-
Physician-applied	ML	RF	0.955 (0.01)	**0.980 (<0.01)**	0.967 (<0.01)	**0.937 (0.01)**	**0.997 (<0.01)**
		XGB	0.951 (0.00)	**0.980 (<0.01)**	0.965 (<0.01)	0.933 (<0.01)	**0.997 (<0.01)**
		SVM	**0.988 (<0.01)**	0.930 (0.00)	0.958 (<0.01)	0.912 (0.01)	**0.997 (<0.01)**
		NN	0.976 (0.01)	0.962 (0.01)	**0.969 (<0.01)**	0.936 (0.01)	**0.997 (<0.01)**
	KL		**0.963 (0.00)**	**0.980 (0.00)**	**0.972 (0.00)**	**0.944 (0.00)**	-

Accordingly, in this iteration, the test set was comprised of one physician- and one self-applied measurement corresponding to the test subject. This iterative process was repeated for a total of 182 times until all 182 subjects who had both physician- and self-applied measurements were tested once.

Four ML models namely RF [17], XGB [18], SVM [19], and NN [20] classifiers were deployed to examine the performance of the ML approach. For each of these four ML models, the process of training and testing the ML classifier was repeated for 10 iterations. Subsequently, the arithmetic mean and standard deviation of sensitivity, specificity, F- measure, Kappa-score, and area under the receiver operating characteristic curve (AUC) were calculated over these 10 iterations and reported for both physician- and self-applied measurements.

For all the four classifiers, we kept the default parameters as no tuning was applied. The parameters for the RF classifier [22] were 1025 trees, Gini as the measure for the quality of splits, square root of the features for the maximum number of features when looking for the best split. For the XGB classifier [18], we used a forest size equal to 1025 trees, step size shrinking (eta) equal to 0.3, minimum loss reduction (gamma) equal to 0, and max depth equal to 6, L_2-regularization (lambda) equal to 1, as well as L_1-regularization (alpha) equal to 0. For the SVM classifier [22], the parameters were the radial basis function for the kernel, (number of features)$^{-1}$ for the kernel coefficient (gamma), and 1 for the penalty term (C). Lastly, for the NN classifier [22], we used a single hidden layer with 100 neurons, rectified linear unit (relu) activation, and an adaptive moment estimation (adam) optimizer.

IV. RESULTS

In this section, the prediction performances of the two approaches, the prediction consistency of the physician- and self-applied data as measured by the KL and ML approaches, as well as the prediction agreement of the KL and ML approaches are presented.

A. PREDICTION PERFORMANCES

The results of the performed classifications on the test set for the KL and ML approaches for both of the physician- and self-applied data are illustrated in Table 2.

1) KL APPROACH RESULTS

For the physician-applied data, the KL approach provided a sensitivity of 0.963 (0.00) and a specificity of 0.980 (0.00). For the self-applied data, the sensitivity and specificity values were 0.976 (0.00) and 0.929 (0.00), respectively. The F- measure was 0.972 (0.00) for the physician-applied data and 0.952 (0.00) for the self-applied data. The Kappa-score of the predicted labels versus true labels was 0.944 (0.00) for the physician-applied data and 0.900 (0.00) for the self-applied data. The KL approach is a deterministic algorithm; therefore, all the standard deviation values are zero. Since the KL algorithm does not provide prediction probabilities, the AUC value was not computed for the KL algorithm results.

2) ML APPROACH RESULTS

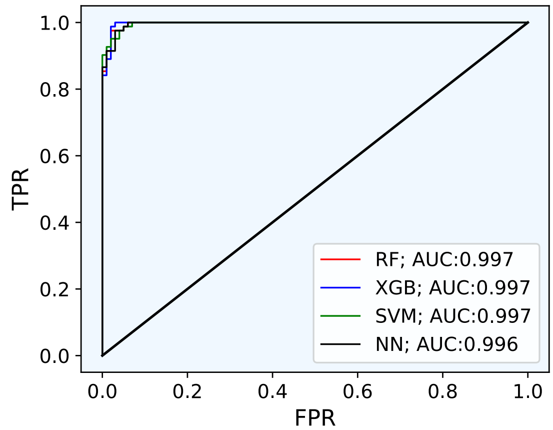
Repeated 10 times, the four RF, XGB, SVM, and NN classifiers were trained and tested following the framework illustrated in Figure 3. The performance metrics calculated for all the four ML classifiers are described in Table 2. According to Table 2, the best sensitivity and specificity values for the physician-applied data were 0.988 ($<$0.01) and 0.980 ($<$0.01) while for the self-applied data they were 0.948 (0.01) and 0.936 (0.01), respectively. The highest acquired F-measure was 0.969 ($<$0.01) for the physician-applied data and 0.937 ($<$0.01) for the self-applied data. The best Kappa-score of the predicted labels versus true labels was 0.937 (0.01) for the physician-applied data and 0.873 (0.01) for the self-applied data. The best AUC for the physician-applied data was 0.997 ($<$0.01) and for the self-applied data was 0.979 (0.00). ROC curves of the predictions made by the ML approach for the physician- and self-applied data are presented in Figure 4a and Figure 4b, respectively.

B. PREDICTION CONSISTENCY OF THE KL APPROACH

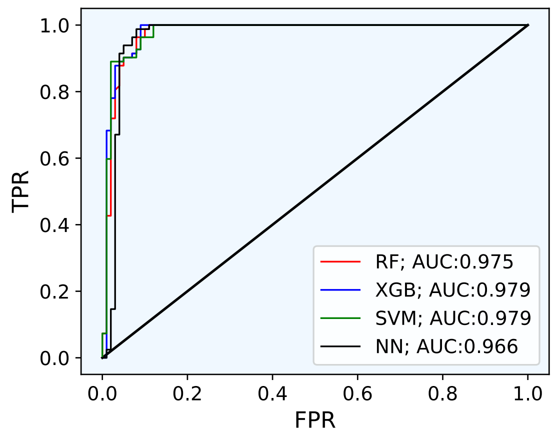
In order for us to compare the exact prediction of each approach for a certain subject, we introduce a barcode plot in which the true label (or the ground truth) for each subject is marked with a color (see Figure 5). In this figure, pink represents the SR category and blue represents the AFib category. Columns C and E in Figure 5 illustrate the predictions made by the KL approach for the physician- and self-applied data, respectively. The agreement of the two measurement scenarios obtained by the KL approach – measured by Kappa-score – was 0.911. This shows a slightly better agreement as compared with the ML approach.

C. PREDICTION CONSISTENCY OF THE ML APPROACH

We sought to compare the performance of the physician- and self-applied AFib detection by selecting the best classifier based on the F-measure values. Accordingly, the NN classifier was chosen as the best classifier as reported in Table 2. The agreement of the predictions made by the two measurement scenarios analyzed by the ML approach was assessed by Kappa-score. The resulting Kappa-score was equal to 0.811 for the physician- versus the self-applied classification.



(a) ROC curves of the four classifiers used for physician-applied data.



(b) ROC curves of the four classifiers used for self-applied data.

FIGURE 4. ROC curves of the ML classifiers for the physician- and self-applied measurements.

A visual representation of the exact predictions made by the NN classifier is presented in columns B and D in Figure 5.

D. PREDICTION AGREEMENT OF THE KL AND THE ML APPROACHES

1) PHYSICIAN-APPLIED MEASUREMENTS

As shown in columns B and C in Figure 5, there were very few misclassifications made by the KL and ML approaches. In the SR group, two and four measurements were wrongly predicted by the KL and ML, respectively. In the AFib group, there were two misclassifications made by the ML approach and three misclassifications made by the KL approach. There was no common misclassification made by both of the ML and KL approaches in the physician-applied data. The agreement of the two approaches for the classification of the physician-applied measurements was equal to 0.900.

2) SELF-APPLIED MEASUREMENTS

According to columns D and E in Figure 5, there were four subjects in the SR group whose measurements were misclassified by both the KL and ML approaches. Furthermore, there were three more misclassifications made by the KL approach

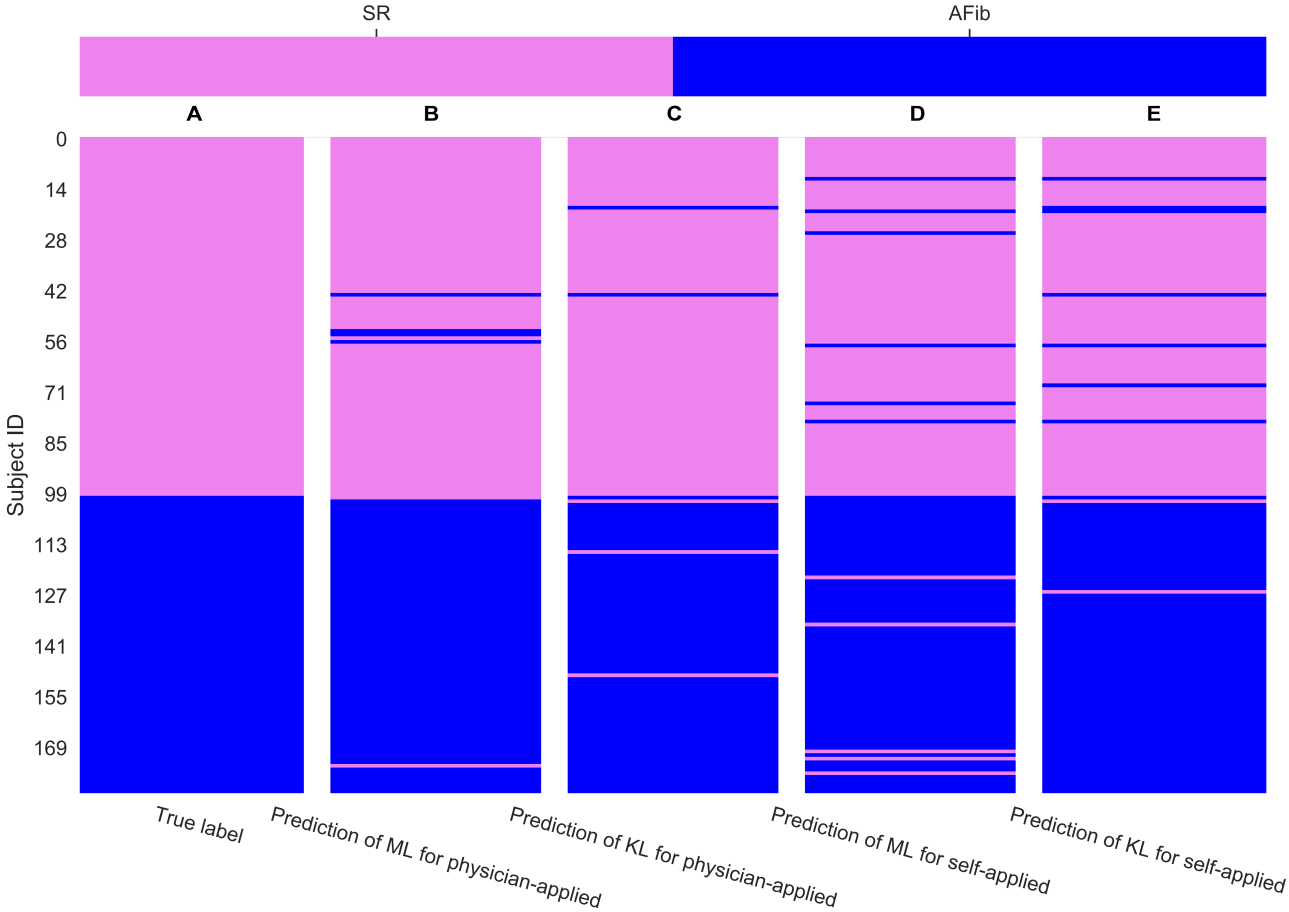


FIGURE 5. **Barcode plot of the predictions made by ML and KL approaches for the physician- (columns B and C) and the self-applied (columns D and E) data in comparison with the true labels (column A). The ground truth of the labels are coded with colors; pink represents the SR category and blue represents the AFib category.**

as well as two more misclassifications made by the ML approach. In the AFib group, there were only two misclassifications made by the KL approach whilst there were five misclassifications made by the ML approach. The two KL and ML approaches did not have any common misclassifications in the AFib group. The agreement of the two approaches for the classification of the self-applied measurements was equal to 0.867.

In Figure 6, a few examples of correctly classified and misclassified measurements are outlined. The presented data in this figure were selected to show the typical style of AFib and SR (subjects 1 and 2) and further clarify the sources of misclassifications (subjects 3 to 5) in the self-applied data. Although occurring only rarely, the sources of error were low-quality data and paroxysmal AFib.

V. DISCUSSION

The subjects who participated in this study were all elderly adults who were admitted to the Department of Cardiology, Turku Hospital, Turku, Finland. Cardiologists recorded the sMCG data from the subjects first and then instructed them to carry out the self-applied measurement. Each subject was given three opportunities to collect at least one successful self-applied measurement. Of the 300 participants, 182 subjects were able and willing to contribute to collecting the self-applied measurement according to the given instructions. The prospective head-to-head analysis was only done for these subjects.

The results of this study are in line with our previous contributions where we introduced the concept of mechanocardiography for AFib detection by considering only physician-applied sMCG recordings [2], [15], [16]. The major contribution of this study was the performance assessment of AFib detection in self-applied sMCG recordings. We demonstrated that self-applied measurements can reliably contribute to the detection of AFib. Additionally, our results are comparable with the latest smartphone-based AFib detection technologies. Recently, research studies have shown that facial pulsatile PPG recorded by a smartphone camera could facilitate AFib detection with sensitivity value of 0.95 and specificity value of 0.96 [23]. Similarly, AFib detection using fingertip PPG collected by a smartphone camera resulted in sensitivity value of 0.92 and specificity value of 1.00 [24]. PPG-based detection of AFib using a

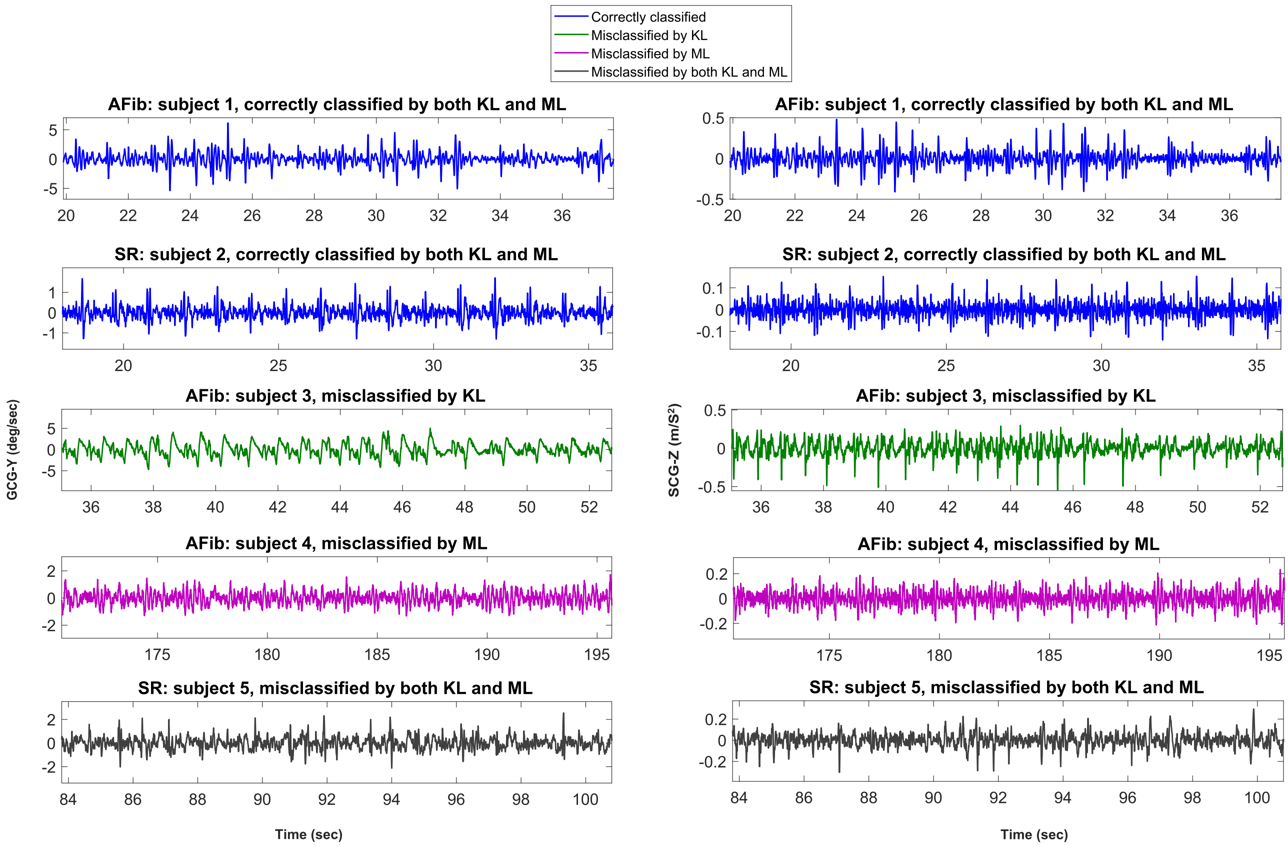


FIGURE 6. From self-applied data, snapshots of correctly classified signals from AFib and SR categories in the two top rows (subject 1 and subject 2), an AFib case misclassified by KL (subject 3), an AFib case misclassified by ML (subject 4), and a SR case misclassified by both KL and ML (subject 5) are provided.

smartwatch provided 0.98 and 0.90 sensitivity and specificity rates [25]. ECG-based detection of AFib using an Apple Watch with a KardiaBand resulted in 0.93 sensitivity as well as 0.84 specificity rates [5].

This study was designed in such a way that those participants who did not have proper self-applied measurements but proper physician-applied measurement were placed into the training set to be exclusively used for training the ML approach. The rest of the subjects who had both physician- and self-applied measurements that were of a proper quality and duration were used as a test set. That is, the testing process was only done for those subjects ($n = 182$) who had both physician- and self-applied measurements.

The barcode plots introduced in this paper facilitated the investigation of the subject-by-subject variations of the predictions made by the ML and KL approaches. The agreement of AFib recognition using the self- and the physician-applied sMCG measurements was investigated in this work. Two different approaches, namely ML and KL, were deployed to assess whether the same level of accuracy can be obtained for the self-applied data in comparison with the physician-applied data. The results of both the ML and KL approaches showed that a slightly lower level for the F-measure and kappa-score were accomplished with the self-applied measurements as compared with the physician-applied measurements.

All the four ML classifiers that were used in this study provided quite similar performance levels in both the physician- and self-applied data sets. This implies that AFib detection was not dependent on the type of the deployed classifiers. The results of the NN classifier, which provided the slightly higher F-measure for both of the physician- and self-applied data, was selected and analyzed further.

According to Table 2, the presented performance metrics for both the ML and KL approaches confirmed that switching to the self-applied data led to negligible performance depreciation. However, the KL approach was better able to tolerate the variations caused by transferring the data acquisition responsibility to the subjects themselves.

In Figure 6, the top two rows (subjects 1 and 2) indicate the typical characteristics of AFib and SR data which are quite straightforward to classify by both the ML and KL approaches. The next row from the top (subject 3) corresponds to an AFib case whose self-applied measurement was misclassified by the KL approach. The signals of this case look mostly regular except for the abrupt transitions from high heart rate to lower heart rate. The fourth row from the top (subject 4) associates with an AFib case whose self-applied measurement was misclassified by the ML approach. It seems the low quality of data confused the ML approach. The bottom row (subject 5) exemplifies a potentially mislabeled measurement as we could barely find any regularity in the data. Such mislabeled measurements were quite rare in our data set. The reason for assigning wrong labels could be the sporadic nature of AFib in some patients.

It is worth noting that all participants in this study were elderly adults who might not have been very competent smartphone users and probably unfamiliar with phones equipped with a touchscreen. Many patients were nervous, in a quite poor condition, and not interested in or capable of concentrating on using a smartphone in this acute hospital setting. Nevertheless, the most common and important reason for an unsuccessful self-applied recording was the unfamiliarity with touch screen devices (maybe sometimes combined with poor eyesight, loss of coordination, physical trauma, etc.). For example, most of the elderly subjects pressed the touch screen very hard, as if they were using a regular phone with a keypad, which resulted in failure of the touch screen to register their touch. It seemed difficult for the elderly patients to learn the right touch intensity; many did not learn; hence no successful recording was made. Those patients who had experience with smartphones, of course, did not have this problem.

On the other hand, the user interface of the deployed signal acquisition application on the mobile phone was not exclusively designed for elderly adults. The smartphone application was primarily meant to be a data logger rather than an application designed for ordinary consumers. Despite all the aforementioned limitations, 182 subjects were willing and able to make a successful self-applied recording. Our results show that by using sMCG we were able to detect AFib with a F-measure value equal to 0.937. The user interface, unfamiliarity with smartphones, and the clinical condition of the subjects while being hospitalized were the major limitations of the self-applied data acquisition in this study.

Another potential source of error causing misclassification, as depicted in Figure 6, could be due to sporadic AFib episodes. This potentially hindering factor must be investigated in future research studies. In the MODE-AF study dataset, the focus was on the discrimination of AFib from SR; however, there are other types of arrhythmias such as atrial flutter, premature atrial contractions, ventricular tachycardia, ventricular fibrillation, and premature ventricular contractions which will be investigated in future studies.

VI. CONCLUSION

Smartphone mechanocardiography is a reliable measurement modality for AFib detection. The findings of this study suggest that in a study sample of 182 elderly adults, both physician- and self-applied measurements could provide sufficient information for successful AFib detection. These findings contribute in several ways to our understanding of the heart mechanical functioning and therefore provide a basis for arrhythmia detection using sMCG. The present study adds to the growing body of research that indicates the practicality of built-in inertial sensors of smartphones and/or wearable devices for continuous monitoring of heart conditions. Further research should be undertaken to explore the effectiveness of sMCG in screening for AFib.

ACKNOWLEDGMENT

The authors would like to thank all the participants in the study.

REFERENCES

[1] P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H.-C. Diener, H. Heidbuchel, J. Hendriks, G. Hindricks, A. S. Manolis, J. Oldgren, B. A. Popescu, U. Schotten, and B. V. Putte, "2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *Eur. J. Cardio-Thoracic Surg.*, vol. 50, no. 5, pp. e1–e88, 2016.

[2] J. Jaakkola, S. Jaakkola, O. Lahdenoja, T. Hurnanen, T. Koivisto, and M. Pänkäälä, T. Knuutila, T. O. Kiviniemi, T. Vasankari, and K. E. 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.

[3] R. W. Rho and R. L. Page, "Asymptomatic atrial fibrillation," *Progr. Cardiovascular Diseases*, vol. 48, no. 2, pp. 79–87, 2005.

[4] J. K. Lau, N. Lowres, L. Neubeck, D. B. Brieger, R. W. Sy, C. D. Galloway, D. E. Albert, and S. B. Freedman, "iPhone ECG application for community screening to detect silent atrial fibrillation: A novel technology to prevent stroke," *Int. J. Cardiol.*, vol. 165, no. 1, pp. 193–194, 2013.

[5] J. M. Bumgarner, C. T. Lambert, A. A. Hussein, D. J. Cantillon, B. Baranowski, K. Wolski, B. D. Lindsay, O. M. Wazni, and K. G. Tarakji, "Smartwatch algorithm for automated detection of atrial fibrillation," *J. Amer. College Cardiol.*, vol. 71, no. 21, pp. 2381–2388, 2018.

[6] M. P. Turakhia, M. Desai, H. Hedlin, A. Rajmane, N. Talati, T. Ferris, S. Desai, D. Nag, M. Patel, P. Kowey, J. S. Rumsfeld, A. M. Russo, M. T. Hills, C. B. Granger, K. W. Mahaffey, and M. V. Perez, "Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The apple heart study," *Amer. Heart J.*, vol. 207, pp. 66–75, Jan. 2019.

[7] T. Hendrikx, M. Rosenqvist, P. Wester, H. Sandström, and R. Hörnsten, "Intermittent short ECG recording is more effective than 24-hour Holter ECG in detection of arrhythmias," *BMC Cardiovascular Disorders*, vol. 14, no. 1, p. 41, 2014.

[8] R. Tieleman, Y. Plantinga, D. Rinkes, G. Bartels, J. Posma, R. Cator, C. Hofman, and R. P. Houben, "Validation and clinical use of a novel diagnostic device for screening of atrial fibrillation," *Europace*, vol. 16, no. 9, pp. 1291–1295, 2014.

[9] P. M. Barrett, R. Komatireddy, S. Haaser, S. Topol, J. Sheard, J. Encinas, A. J. Fought, and E. J. Topol, "Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring," *Amer. J. Med.*, vol. 127, no. 1, pp. 95.e11–95.e17, Jan. 2014.

[10] C. Mortelmans, "Validation of a new smartphone application ('FibriCheck') for the diagnosis of atrial fibrillation in primary care," KU Leuven, Leuven, Belgium, Tech. Rep., 2016.

[11] L. Krivoshei, S. Weber, T. Burkard, A. Maseli, N. Brasier, M. Kühne, D. Conen, T. Huebner, A. Seeck, and J. Eckstein, "Smart detection of atrial fibrillation," *Europace*, vol. 19, no. 5, pp. 753–757, 2016.

[12] S. Nemati, M. M. Ghassemi, V. Ambai, N. Isakadze, O. Levantsevych, A. Shah, and G. D. Clifford, "Monitoring and detecting atrial fibrillation using wearable technology," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 3394–3397.

[13] A. J. Camm, P. Kirchhof, G. Y. Lip, and E. A. Schotten, "Guidelines for the management of atrial fibrillation: The task force for the management of atrial fibrillation of the European society of cardiology (ESC)," *Eur. Heart J.*, vol. 31, no. 19, pp. 2369–2429, 2010.

[14] V. Fuster *et al.*, "ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: Executive summary: A report of the American college of cardiology/American heart association task force on practice guidelines and the European society of cardiology committee for practice guidelines and policy conferences (committee to develop guidelines for the management of patients with atrial fibrillation) developed in collaboration with the north American society of pacing and electrophysiology," *J. Amer. College Cardiol.*, vol. 38, no. 4, pp. 1231–1265, 2001.

[15] O. Lahdenoja, T. Hurnanen, Z. Iftikhar, S. Nieminen, T. Knuutila, A. Saraste, T. Kiviniemi, T. Vasankari, J. Airaksinen, M. Pänkäälä, and T. Koivisto, "Atrial fibrillation detection via accelerometer and gyroscope of a smartphone," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 108–118, Jan. 2018.

[16] M. J. Tadi, S. Mehrang, M. Kaisti, O. Lahdenoja, T. Hurnanen, J. Jaakkola, S. Jaakkola, T. Vasankari, T. Kiviniemi, J. Airaksinen, T. Knuutila, E. Lehtonen, T. Koivisto, and M. Pänkäälä, "Comprehensive analysis of cardiogenic vibrations for automated detection of atrial fibrillation using smartphone mechanocardiograms," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2230–2242, Nov. 2018.

[17] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.

[21] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Springer, 2013.

[22] 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 É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[23] B. P. Yan, W. H. Lai, C. K. Chan, S. C.-H. Chan, L.-H. Chan, K.-M. Lam, H.-W. Lau, C.-M. Ng, L.-Y. Tai, K.-W. Yip, O. T. L. To, B. Freedman, Y. C. Poh, and M.-Z. Poh, "Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals," *J. Amer. Heart Assoc.*, vol. 7, no. 8, 2018, Art. no. e008585.

[24] N. Brasier, C. J. Raichle, M. Dörr, A. Becke, V. Nohturfft, S. Weber, F. Bulacher, L. Salomon, T. Noah, R. Birkemeyer, and J. Eckstein, "Detection of atrial fibrillation with a smartphone camera: First prospective, international, two-centre, clinical validation study (DETECT AF PRO)," *EP Europace*, vol. 21, no. 1, pp. 41–47, 2018.

[25] G. H. Tison, J. M. Sanchez, B. Ballinger, A. Singh, J. E. Olgin, M. J. Pletcher, E. Vittinghoff, E. S. Lee, S. M. Fan, and R. A. Gladstone, C. Mikell, N. Sohoni, J. Hsieh, and G. M. Marcus, "Passive detection of atrial fibrillation using a commercially available smartwatch," *JAMA Cardiol.*, vol. 3, no. 5, pp. 409–416, 2018.

SAEED MEHRANG was born in Isfahan, Iran, in 1990. He received the M.Sc. degree in information technology from the Tampere University of Technology, in 2016. He is currently pursuing the Ph.D. degree with the Health Technology Group, Faculty of Science and Engineering, University of Turku. He has cooperated as a Data Scientist in several research projects. His research is mainly focused on the applications of deep learning, machine learning, and signal and image processing for medical and healthcare technologies.

MOJTABA JAFARI TADI was born in Isfahan, Iran, in 1989. He received the B.Sc. degree in biomedical engineering, in 2012, the M.Sc. degree in biomedical imaging from Åbo Academi University, and the Ph.D. degree in medical physics and engineering from the University of Turku, Finland, in 2014 and 2018, respectively, where he has been a Senior Researcher/Postdoctoral Fellow with the Health Technology Group, Faculty of Science and Engineering, since 2019. His research interests include noninvasive physiologic monitoring for human health and medical imaging, and developing signal processing and machine learning techniques for detecting chronic diseases.

TERO HURNANEN received the master's degree from the Department of Physics, University of Turku, in 2006, where he is currently a Researcher with the Department of Future Technology. He has been previously working in areas of spectroscopy, digital signal processing, and telecommunication. He is currently researching algorithms for automatic detection of atrial fibrillation.

TIMO KNUUTILA is currently an Adjunct Professor with the Department of Future Technologies, University of Turku. He has excellent knowledge of data mining, algorithms, and machine learning techniques.

OLLI LAHDENOJA received the M.Sc. and D.Sc. (Tech) degrees from the University of Turku (UTU), Finland, in 2003 and 2015, respectively, where he is currently a Senior Researcher with the Department of Future Technologies, within the area of biomedical engineering. His research interests include biomedical engineering, biomedical signal processing, and computer vision. He has worked in several research projects related to a concrete implementation of systems related to these areas. He has published several peer-reviewed international journals and conference papers in these fields.

JUSSI JAAKKOLA was born in Pirkkala, Finland, in 1991. He received the Medicinae Doctor (M.D.) degree and the Ph.D. degree in cardiology and cardiovascular medicine from the University of Turku, Finland, in 2016 and 2018, respectively. He is currently an Internal Medicine Resident with the Satakunta Central Hospital (Satasairaala), Pori, Finland. His current research interests include the screening of atrial fibrillation and aortic valve replacement therapy.

SAMULI JAAKKOLA was born in 1981. He received the university degree (LL), the Medicinae Doctor (M.D.) degree, and the Ph.D. degree in cardiology and cardiovascular medicine from the University of Turku, Finland, in 2010, 2016, and 2018, respectively. He is currently a Cardiology Specialist with the Turku Heart Center, Turku, Finland.

TUIJA VASANKARI is a Registered Nurse and is currently a Study Coordinator with the Heart Center of Turku University Hospital. Her expertise includes study protocols, patient information and informed consent forms, and clinical trials.

TUOMAS KIVINIEMI was born in Finland, in 1979. He received the M.D. degree, in 2006, the Ph.D. degree, in 2006, and the Adjunct Professorship, in 2013, from the University of Turku, Finland, where he has been a Specialist in cardiology, since 2012. As a Clinical Cardiologist, Clinical Trialist, and Translational Researcher with a special research interest in atrial fibrillation (AF) and long-term complications after percutaneous and surgical procedures, he has produced over 90 peer-reviewed publications in English. He has been a Visiting Scientist with Brigham and Women's Hospital, Harvard Medical School, Boston, USA, since 2018. His current research interests include clinical trials in the field of cardiovascular medicine, detection of AF, and translational research in the field of cell biology of AF and atrial cardiomyopathy. He served as the Secretary and the President, and serves as the Past-President for EAPCI Working Group of Finnish Cardiac Society. He has been a Fellow of European Society of Cardiology (FESC), since 2017.

JUHANI AIRAKSINEN is currently a Professor in cardiology and the Director of the Heart Center, Turku University Hospital. He has supervised 13 doctoral theses and has published over 400 peer-reviewed articles. His main research interests include atrial fibrillation and antithrombotic treatment.

TERO KOIVISTO received the M.Sc. degree in system circuits from the University of Turku, in 2004. He has approximately 15 invention disclosures and five patents pending. His research interests include integrated circuits for ultra-low-power autonomous biomedical sensors and cardiac measurement techniques, especially, atrial fibrillation detection using MEMS accelerometers and pressure sensors. Since 2011, he has managed several projects in the field of cardiac monitoring and is currently operating as a Research and Development Leader in these projects.

MIKKO PÄNKÄÄLÄ received the master's, licentiate's, and doctor's degrees in microelectronics from the University of Turku, in 2004, 2008, and 2014, respectively, where he is currently an Adjunct Professor with the Department of Future Technologies. He has extensive expertise in biosignal processing. He has been involved with the research projects cardiac monitoring via accelerometer and gyroscope sensors. He has also had a key role in the development methods and algorithms for an accelerometer-based detection of atrial fibrillation. He has published 31 peer-reviewed publications and has one granted and six pending patents regarding cardiac measurements, apparatus, and algorithms.