Sensor applications

Evaluation of Capacitive ECG for Unobtrusive Atrial Fibrillation Monitoring

Winston Zhang[1], Zhi Li[1], Jonathan Gryak[2], Pujitha Gunaratne[3], Emily Wittrup[1], and Kayvan Najarian[1,4,5,6*]

[1]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109 USA
[2]Department of Computer Science, Queens College, City University of New York, New York, NY 11367 USA
[3]Collaborative Safety Research Center, Toyota Motor Engineering & Manufacturing North America, Inc., Ann Arbor, MI 48109 USA
[4]Michigan Institute for Data Science, University of Michigan, Ann Arbor, MI 48109 USA
[5]Department of Emergency Medicine, University of Michigan, Ann Arbor, MI 48109 USA
[6]Max Harry Weil Institute for Critical Care Research and Innovation, University of Michigan, Ann Arbor, MI 48109 USA
*Senior Member, IEEE

Manuscript received 14 June 2023; revised 5 September 2023; accepted 8 September 2023. Date of publication 14 September 2023; date of current version 2 October 2023.

Abstract—Unobtrusive collection of vital signs using sensors embedded in beds, chairs, and automobile seats can longitudinally monitor patients for abnormal heart conditions outside of the hospital to inform both preventative and postdiagnosis care. The capacitive electrocardiogram (cECG) shows potential for collecting electrical information about a patient's heart without requiring skin contact like regular electrocardiogram (ECG). However, motion artifacts and environmental factors easily corrupt cECG signal quality and reduce the diagnostic value of unobtrusively collected cECG. To evaluate the atrial fibrillation (AF) screening performance of cECG compared to ECG, we conduct three different experiments on the clinical UnoViS dataset consisting of 55 min of concurrent ECG and cECG signals from 92 patients with manual clinician annotations of AF. First, we trained and evaluated models to detect AF events on cECG and ECG separately. Then, we trained a model using ECG and evaluated it on cECG to measure the interchangeability of the ECG and cECG domains. For each experiment five-fold subjectwise, class-stratified cross-validation was used to assess trained algorithm performance on hold-out test sets and three different algorithmic methodologies were assessed. Although evaluating the trained ECG model on cECG data (AUC: 0.874 $\pm$ 0.067, $F1$-score: 0.553 $\pm$ 0.148) performed slightly worse than evaluating on ECG, it did perform better than the model trained and evaluated on cECG alone, which suggests that utilizing ECG data for model training can effectively screen similar conditions in cECG data.

Index Terms—Sensor applications, capacitive electrocardiogram (cECG), atrial fibrillation (AF), noninvasive sensors.

I. INTRODUCTION

Heart rhythm disorders, such as arrhythmia, were listed as a cause of death for more than half a million people in 2018 [1], and it is projected that 12.1 million people may have atrial fibrillation (AF) by 2030 [2]. Early detection and treatment of AF is critical to improving clinical outcomes [3]. Specifically, patients with incidentally detected asymptomatic AF (AF detected after screening) have increased risk of mortality due to cardiovascular and all-cause mortality compared with patients with typical AF symptoms [4].

Heart rhythm disorders are routinely detected via an electrocardiogram (ECG), which records the difference in electric potential between electrodes placed on the skin surface. The electric signals generated by the atrial sine node and propagated through the heart muscles are then picked up by the electrodes and converted to a digital signal [5]. When abnormalities occur in the electrical pulses of the heart, such as during AF, corresponding changes can be observed in recorded ECG signals and used by clinicians for diagnosis.

However, long-term screening for arrhythmia requires nonintrusive sensors that can be incorporated into a patient's life more comfortably [6]. First introduced in 1969 [7], the capacitive ECG (cECG) allows for measurement of electrical signals without requiring skin contact, making cECG well-suited for long-term monitoring of ECG through electrodes embedded in everyday objects, such as car seats and beds [8]. Previous work, such as Weil et al. [9], have concluded that cECG signals could identify ST-elevation myocardial infarction in patients. Bhardwaj et al. found coherence between measured cECG, and ECG signals collected from sensors embedded in a car seat during driving.

Despite the convenience of noncontact electrodes for long-term monitoring, cECG are more susceptible to motion and static electrical artifacts when compared with skin contact ECG [8]. Previous work by Czaplik et al. [10] investigated the discrepancy between clinician annotations of cECG and ECG signals recorded in the seated position. Although high correspondence was achieved for clinician labeled AF between signal modalities, a significant number of collected cECG signals were thrown out due to artifacts from environmental factors.

Recent work has attempted to overcome the volatile nature of the cECG modality through methods, such as denoising [11] and signal fusion [12]. Current cECG-based comparisons are also limited by the dearth of publicly available unobtrusive signal databases as compared with large clinical ECG databases, such as Physionet [13] and MIT-BIH [14]. A potential solution to this lack of data is to train diagnostic algorithms on the cleaner and abundant ECG modality to learn more accurate representations of cardiac conditions and then applying the models for diagnosis of cECG signals.

In this letter, we will compare the AF detection performance of models trained and evaluated on a single modality (ECG and cECG)

Corresponding author: Emily Wittrup (e-mail: ewittrup@umich.edu).
Associate Editor: K. Ozanyan.
Digital Object Identifier 10.1109/LSENS.2023.3315223

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

TABLE 1. Cohort Size After Removing Uninformative Signals

Cohort	Patients (no.)	5 s intervals (no.)
Atrial fibrillation	13	59
Control	77	410
Total	90	469

with the evaluation performance of a model trained on ECG and evaluated on cECG data.

II. DATASET

The UnoViS dataset contains unobtrusive signals, including photoplethysmogram, ECG, and cECG, collected from diverse scenarios, such as in driving, bedside, and clinical environments [15]. In this letter, we utilized the "UnoViS clin2009" dataset, which contains 55 min of single lead cECG and ECG recordings simultaneously collected from 92 patients while seated along with AF annotations from two clinicians [10]. ECG was recorded at 500 Hz sampling rate, while cECG was recorded at 125 Hz sampling rate. Patient age was in the range 64.3 ± 21.9 years with an average data recording time per patient of 35.89 s. Details about clinical design, sensor make-up, and environmental controls can be found in [10] and [15].

The raw ECG and cECG signals were preprocessed by first up sampling the cECG signals to 500 Hz to match the sampling rate of the ECG recordings. Then, a fourth-order bandpass Butterworth filter with cutoffs at 0.5 and 40 Hz was applied to both the ECG and cECG signals to remove noise. Baseline wander was then removed using a double median filter at orders 0.2 and 0.6 times the sampling frequency.

We divided each patient's recordings into 5-s intervals of corresponding cECG and ECG signals. Each section was normalized to be in the range $(-1, 1)$. Certain signal sections within the dataset were manually observed to be flat signals or had no observable R peaks. We removed uninformative signal intervals that contained more than 3 s of flat signal or less than two annotated R peaks, resulting in complete removal of signals from two patients. The resulting number of patients and 5 s intervals is shown in Table 1.

III. METHODS

Three diagnostic algorithms for AF were implemented and compared with each other on this dataset which will be discussed in more detail in the following sections: a probabilistic deterministic finite-state automata (PDFA) based algorithm [16], a convolutional neural network and long short-term memory network (CNN–LSTM) deep-learning model [17], and a random forest (RF) classifier using heart rate variability (HRV) features extracted as in Asl et al. [18].

We formulated AF detection as a binary classification problem and trained and evaluated the algorithms using the proposed methodology in Fig. 1. For each algorithm, we trained on the dataset and evaluated on a hold-out test set consisting of the other of the dataset. During training, we used subjectwise class-stratified five-fold cross-validation (CV) where patients are randomly sorted into folds to ensure a balanced class ratio between folds. A hyperparameter grid search is performed over the five-fold CV, and the best combination of parameters is selected based on averaged validation $F1$-score. We then use an ensemble of the five trained models from the best CV for evaluation on the test set. This procedure is replicated 10 times over 10 random train–test splits to reduce the impact that random splitting has on performance results.

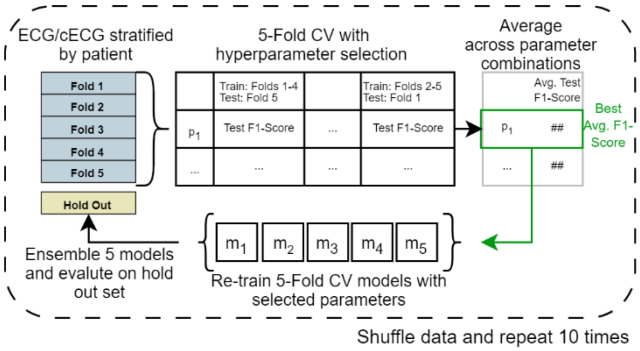

Fig. 1. Proposed training and evaluation procedure. CV = cross-validation.

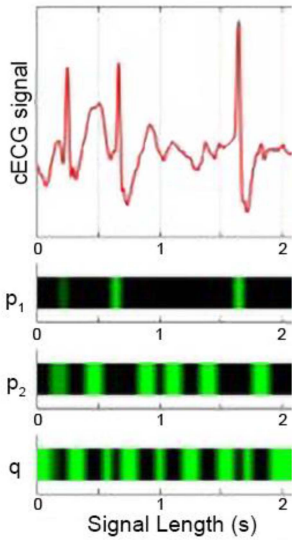

Fig. 2. Sample cECG from atrial fibrillation patient encoded into a ternary alphabet {$p1$, $p2$,q}. Green represents high probability for the symbol at specified point in time.

Final reported results are the mean ± standard deviation (SD) test results across all 10 repetitions. This process was designed to address the potential shortcomings of utilizing a relatively small dataset in terms of predictive power and generalizability.

A. PDFA-Based Algorithm

The PDFA-based algorithm is a finite-state automata that stochastically generates sequences of letters from a fixed alphabet [19]. Each state has a transition probability for each unique letter in the alphabet, which determines the next state. The inputs of the PDFA model for arrhythmia prediction by Li et al. [16] first encodes ECG signals into a probabilistic string with symbols roughly corresponding to important subsections of a QRS complex, as can be seen in Fig. 2. A separate frequency prefix tree (FPT) of observed combinations of symbols is then constructed using the training set of encoded signals for both AF and control cases. After pruning the FPTs based on a tuned cutoff frequency parameter C, new signals are classified through expected likelihood comparison between the two PDFA distributions. In this letter, we grid search over the parameter C for both AF and control PDFAs using suggested ranges from [16].

TABLE 2. Performance Metrics for Each Algorithm and Experiment Across 10 Repetitions Presented as Mean ± Standard Deviation

Experiment	Algorithm	AUC	F1	Sensitivity	Specificity	Precision
1	PDFA	0.760 ± 0.137	0.459 ± 0.122	0.714 ± 0.165	0.776 ± 0.080	0.355 ± 0.118
	CNN+LSTM	0.785 ± 0.098	0.444 ± 0.118	0.813 ± 0.060	0.702 ± 0.134	0.334 ± 0.126
	HRV+RF	**0.885 ± 0.084**	**0.583 ± 0.140**	0.836 ± 0.047	0.834 ± 0.105	0.467 ± 0.163
2	PDFA	0.688 ± 0.102	0.366 ± 0.086	0.678 ± 0.163	0.661 ± 0.161	0.275 ± 0.082
	CNN+LSTM	0.606 ± 0.166	0.330 ± 0.131	0.739 ± 0.096	0.525 ± 0.195	0.243 ± 0.129
	HRV+RF	**0.873 ± 0.100**	**0.542 ± 0.189**	0.846 ± 0.046	0.762 ± 0.167	0.445 ± 0.234
3	PDFA	0.782 ± 0.102	0.447 ± 0.101	0.762 ± 0.094	0.720 ± 0.120	0.339 ± 0.116
	CNN+LSTM	0.582 ± 0.111	0.307 ± 0.075	0.768 ± 0.104	0.486 ± 0.116	0.214 ± 0.066
	HRV+RF	**0.874 ± 0.067**	**0.553 ± 0.148**	0.797 ± 0.073	0.823 ± 0.095	0.451 ± 0.176

AUC = area under the receiver operator curve, PDFA = probabilistic deterministic finite-state automata, CNN = convolutional neural network, LSTM = long short-term Memory, HRV = heart rate variability, RF = random forest.

B. CNN–LSTM Deep-Learning Algorithm

For each 5-s ECG and cECG interval, a spectrogram is computed using a Tukey window of 64, shape parameter 0.25, and 50% overlap. A log transform is then computed over the spectrogram. For the model architecture, we used the CRNN proposed by Zihlmann et al. [17] composed of blocks of six consecutive Conv2D, batch normalization, and ReLU activation layers. Parameters selection and details on dropout burst and random resampling data augmentation methods follow the procedures in [17]. The final layer of the model is changed to a dense layer with 1 output using sigmoid activation to fit the binary classification problem. Loss weights for model training are set to the ratio between AF and control samples in the training set.

No grid search was conducted for the CNN–LSTM model due to time limitations. During five-fold CV, the validation set is used for early stopping and learning rate scheduling. The five trained CNN–LSTM models from CV are evaluated on the hold-out set as an ensemble.

C. HRV Features + RF Algorithm

For each 5-s signal in the dataset, the Pan–Tompkins algorithm [20] is used for automatic annotation of R peak locations. A set of 17 unique HRV features are then extracted based on the R peak locations. Besides the seven linear time-domain and seven nonlinear features as listed in [18], three additional linear frequency features are added: the individual high-frequency (HF), the low-frequency (LF) values, and the LF/HF ratio. Extracted features are fed into the RF classifier implemented in the scikit-learn package in Python [21]. The classifier is set to default hyperparameters except for number of trees (200) and number of candidate variables considered each split (grid searched in the range 1–17).

IV. EXPERIMENTS AND RESULTS

In Experiment 1 and Experiment 2, we compare the binary classification performance of ECG versus cECG across the three AF diagnosis algorithms, training and testing on a single data modality. In Experiment 3, we trained on ECG and evaluated performance on the cECG modality.

Significance tests between each pair of models in each experiment are performed using a paired student's t-test on the results distributions from 10 repeated CVs. The t-test is a corrected resampled t-test first proposed by Nadeau and Bengio [22] that reduces Type-I error. We adjust for multiple significance tests using the Benjamini Hochberg procedure [23]. The same significance test procedure is applied to all performance metrics.

A. Experiment 1: Train and Evaluate on ECG

As shown in Table 2, compared to PDFA and CNN + LSTM models, the HRV + RF algorithm performs significantly better across all metrics in AUC (0.885 ± 0.084) and $F1$-score (0.583 ± 0.140). HRV + RF performs better in all other metrics; however, the improvement is not statistically significant. Performance across 10 random splits shows high SD for all three compared diagnostic algorithms, specifically in $F1$-score.

B. Experiment 2: Train and Evaluate on cECG

The HRV + RF algorithm outperforms both other models in all metrics, with a significant increase in AUC (Table 2). The switch to cECG modality resulted in a performance drop for all models in Experiment 2 compared to Experiment 1. Specifically, HRV + RF had a 0.04 drop in mean $F1$-score and an overall increase in performance instability, measured by SD, across repetitions when compared with Experiment 1.

C. Experiment 3: Train on ECG and Evaluate on cECG

The HRV + RF model still outperforms both other models across all metrics and has significantly better mean AUC and $F1$-score, 0.874 ± 0.067 and 0.553 ± 0.148, respectively (Table 2). Compared with Experiment 2, HRV + RF had a slight increase in most performance metrics and a general decrease in SD. Performance is still lower than Experiment 1, however, with only a 0.03 gap in mean $F1$-score and a 0.016 gap in mean precision. PDFA mean performance for all metrics also improved in Experiment 3 over Experiment 2. CNN + LSTM in Experiment 3 performed worse than Experiment 2 in contrast to the other compared models, with a decrease in mean $F1$-score and precision.

V. DISCUSSION

The highly imbalanced nature of the dataset (59 AF intervals compared to 410 control) means that the $F1$-score and precision metrics serve as best comparison between experiments and models.

From results across all three evaluated models, training and testing on the cECG modality in Experiment 2 results in the lowest AF detection performance. This matches with the consensus in literature that cECG signals are more susceptible to noise and do not contain as much information as traditional ECG.

When comparing the Experiments 1 and 2, the HRV + RF model only suffered a drop of 0.04 in $F1$-score compared with larger drops

in $F1$-score by PDFA and CNN + LSTM. From Table 2, the HRV + RF model remained relatively invariant to data modality changes compared to the other two models, with a mean $F1$-score around 0.55. This may be due to the inherent robustness of the HRV + RF model feature format to noise. The loss in information and increased noise in cECG adversely affects models that depend on the full cECG time series for analysis. For example, the PDFA model encodes the full 5 s sequence into symbolic labels and learns to classify AF using characteristics of the QRS waveform. Likewise, the CNN + LSTM uses log spectrograms of the 5-s sequence as input. HRV features, however, only depend on accurate R peak detection. Signal artifacts in cECG have less of an impact on R peak detection compared to more sensitive wave morphology, such as P and T waves.

Performance for PDFA and HRV + RF improved from Experiment 2 to Experiment 3, where ECG data were substituted for cECG during training, suggesting that the relative noisiness of cECG compared to ECG makes it harder for algorithms to learn how to diagnose AF. In contrast, CNN + LSTM performance decreased slightly, perhaps due to the model complexity, which makes it prone to overfitting, specifically on such a small dataset.

VI. CONCLUSION

In conclusion, we observed that simpler algorithms detecting AF from cECG signals benefited from training with the higher quality ECG compared to cECG which is prone to noise and artifacts. This suggests that models trained on large ECG databases can learn informative features for the diagnosis of AF which can be applied to noninvasive cECG signals for longitudinal patient monitoring of heart rhythm disorders. Future work would ideally address a limitation of this letter by expanding to a larger, balanced dataset with signals from patients in a less controlled setting performing a variety of activities. However, the results of this letter demonstrate the utility of transfer learning for signal-based event predictions. Potential applications that may benefit from this approach include detecting cardiac events with wearable devices or objects, such as beds or car seats into which noninvasive sensors can be embedded.

ACKNOWLEDGMENT

This work was supported by the Toyota Motor Engineering & Manufacturing North America, Inc.

REFERENCES

[1] E. J. Benjamin et al., "Heart disease and stroke statistics—2018 update: A report from the American Heart Association," *Circulation*, vol. 137, no. 12, pp. e67–e492, 2018.

[2] S. Colilla, A. Crow, W. Petkun, D. E. Singer, T. Simon, and X. Liu, "Estimates of current and future incidence and prevalence of atrial fibrillation in the us adult population," *Amer. J. Cardiol.*, vol. 112, no. 8, pp. 1142–1147, 2013.

[3] B. Freedman et al., "Screening for atrial fibrillation: A report of the AF-SCREEN international collaboration," *Circulation*, vol. 135, no. 19, pp. 1851–1867, 2017.

[4] K. C. Siontis et al., "Typical, atypical, and asymptomatic presentations of new-onset atrial fibrillation in the community: Characteristics and prognostic implications," *Heart Rhythm*, vol. 13, no. 7, pp. 1418–1424, 2016.

[5] E. J. D. S. Luz, W. R. Schwartz, G. Camara-Ch´ avez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Prog. Biomed.*, vol. 127, pp. 144–164, 2016.

[6] L. Leicht et al., "Capacitive ECG monitoring in cardiac patients during simulated driving," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 3, pp. 749–758, Mar. 2019.

[7] A. Lopez and P. C. Richardson, "Capacitive electrocardiographic and bioelectric electrodes," *IEEE Trans. Biomed. Eng.*, vol. BME-16, no. 1, pp. 99–99, Jan. 1969.

[8] Y. Sun and X. B. Yu, "Capacitive biopotential measurement for electrophysiological signal acquisition: A review," *IEEE Sensors J.*, vol. 16, no. 9, pp. 2832–2853, May 2016.

[9] M. B. Weil, M. Oehler, M. Schilling, and L. S. Maier, "First clinical evaluation of a novel capacitive ECG system in patients with acute myocardial infarction," *Clin. Res. Cardiol.*, vol. 101, no. 3, pp. 165–174, 2012.

[10] M. Czaplik et al., "The reliability and accuracy of a noncontact electrocardiograph system for screening purposes," *Anesth. Analg.*, vol. 114, no. 2, pp. 322–327, 2012.

[11] V. Ravichandran et al., "Deep network for capacitive ECG denoising," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, 2019, pp. 1–6.

[12] C. H. Antink, E. Breuer, D. U. Uguz, and S. Leonhardt, "Signal-level fusion with convolutional neural networks for capacitively coupled ECG in the car," in *Proc. Comput. Cardiol. Conf.*, 2018, pp. 1–4.

[13] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[14] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.

[15] T. Wartzek, M. Czaplik, C. H. Antink, B. Eilebrecht, R. Walocha, and S. Leonhardt, "UnoVIS: The MediT public unobtrusive vital signs database," *Health Inf. Sci. Syst.*, vol. 3, no. 1, pp. 1–9, 2015.

[16] Z. Li et al., "Prediction of cardiac arrhythmia using deterministic probabilistic finite-state automata," *Biomed. Signal Process. Control*, vol. 63, 2021, Art. no. 102200.

[17] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *Proc. Comput. Cardiol.*, 2017, pp. 1–4.

[18] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artif. Intell. Med.*, vol. 44, no. 1, pp. 51–64, 2008.

[19] A. Clark and F. Thollard, "Pac-learnability of probabilistic deterministic finite state automata," *J. Mach. Learn. Res.*, vol. 5, pp. 473–497, 2004.

[20] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[21] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[22] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Adv. Neural Inf. Process. Syst.*, vol. 12, 1999.

[23] D. Thissen, L. Steinberg, and D. Kuang, "Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons," *J. Educ. Behav. Statist.*, vol. 27, no. 1, pp. 77–83, 2002.