Congestive Heart Failure Detection by Short-term Heart Rate Variability for Fast Reference Advice in Urgent Medical Conditions

Yi Zhang[1,2,3], Yuning Qiao[1,5],Peng Xu[2,3], Steven Su[4], Limei Xu[1],Dezhong Yao[2,3]

Abstract— **This study proposed the detection approach for the Congestive Heart Failure (CHF) disease by short-time electrocardiographic monitoring. Recent literature reviews only reported that RR intervals and Heart Rate Variability (HRV) indicate key hidden information to discriminate CHF groups from healthy controls. However what if possible to find certain short-time electrocardiographic monitoring duration to give fast reference advice for CHF diagnoses, has not been well addressed. In this study, commonly applied databases from PhysioNet are introduced, in which approximate 20-hour individual Electrocardiogram (ECG) recordings are used. Those signals are first classified into largely variable assessment lengths. Based on R-R intervals in the assessment length, raw R-R intervals, mean and standard deviation (STD) of R-R intervals, and clinically standard features of short-term (5-min) Heart Rate Variability (HRV), are comparatively analyzed, while combining with classifiers of Recurrent Neural Network (RNN), Random Forest (RF), and Support Vector Machine (SVM). The Leave-one-out Cross-Validation (LOOCV) is adopted for performance verifications, by which the model extracted from certain assessment length will be utilized to test measured data of a subject with the same length. Results show that based on the testing database, a specific 30-minute duration can be achieved by choosing HRV features in full. It is believed that a short-time electrocardiographic monitoring for the CHF detection could be feasible if the standard HRV features together with the classifier of RF or RNN are adopted. It implies that a short-time electrocardiographic monitoring can be applied for fast reference advice of CHF in urgent medical conditions.**

I. INTRODUCTION

The Congestive Heart Failure (CHF) is the condition of the heart that reduces the volume of the cardiac output and slows down the blood flowing rate, and thus a timely diagnosis will definitely be required [1]. In clinical activities the CHF diagnosis is routinely concerned by the report outcomes collected from Electrocardiogram (ECG), B-type Natriuretic Peptide blood test, Echocardiography, and chest radiography [2]. Patients need to physically attend or be delivered through the ambulance system to hospitals or medical centers, and wait time for medical appointments.

[1]The School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, 611731, China

[2]Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 611731, China

[3]Center for Information in BioMedicine, University of Electronic Science and Technology of China, Chengdu, 611731, China yi.zhang@uestc.edu.cn, xupeng@uestc.edu.cn, dyao@uestc.edu.cn

[4]Centre for Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2220, Australia steven.su@uts.edu.au

[5]The School of Engineering and Information Technologies, University of Sydney, Sydney, NSW 2050, Australia yqia2711@uni.sydney.edu.au

Indeed, ECG signals can noninvasively monitor electrical activities of the human heart. Moreover the ECG derivative information such as Heart Rate Variability (HRV) was clinically confirmed to have the potential to heart failures [3]. Therefore, it is naturally substantial to develop an automated diagnostic system for the CHF detection by monitoring ECG signals. Then the analysing result will be presented to medical experts as fast reference advice in urgent conditions in either hospital or free-living environments for considerations of their diagnostic decision.

Many studies dedicated to find robust approaches for the CHF detection. Public clinical data from Beth Israel Deaconess Medical Center (BIDMC) CHF database and Normal Sinus Rhythm R-R Interval Database (NSRID) were popularly used for outcome comparisons [4]. The most recent studies from Liu et al. introduced three non-standard features in terms of time, frequency, and nonlinear behaviours of short-term (5-min) HRV measures to differ CHF patients from NSR controls. They validated the classification performance by Support Vector Machine (SVM) and achieved 100% for both sensitivity and specificity [5]. Another significant outcome was reported by Masetic and Subasi, who used features extracted from the autoregressive Burg method, and examined accuracy values by using classifiers C4.5 Decision Tree (C4.5 DT), Random Forest (RF), k-KNN, Artificial Neural Network (ANN) and SVM. They found the RF algorithm reached the highest success rate of 100% for CHF classifications [6]. Those evidence concluded that both ECG and HRV are key significant features for the CHF detection. Based on the k-fold cross validation, the simply but robust classifiers, SVM and RF, can provide reliable discriminative outcomes (v.s. other classifier tools). However, as far as authors' knowledge the follow-up survey for what if possible to find certain short-time electrocardiographic monitoring duration to give fast reference advice for CHF diagnoses, has not been reported.

This study extends the survey of CHF detection methods to limited electrocardiographic monitoring durations. The goal of this study is to provide fast reference advice for a definite CHF diagnosis in urgent medical conditions. Based on the clinical data, ECG signals of each individual are first segmented into largely variable assessment lengths, each representing the specific electrocardiographic monitoring duration. A 50%-overlapping segment was taken between neighboring assessment lengths. Based on those data in the assessment length, raw R-R intervals, mean and standard deviation

(STD) of R-R intervals, short-term (5-min) HRV in part, and short-term (5-min) HRV in full are comparatively analyzed, while combining with classifiers, Recurrent Neural Network (RNN), RF, and SVM. The Leave-one-out Cross-Validation (LOOCV) is considered for performance verifications.

II. EXPERIMENTS

The databases used in this study was publicly available from PhysioNet, including Beth Israel Deaconess Medical Center (BIDMC) CHF database and Normal Sinus Rhythm R-R Interval Database (NSRID). The BIDMC CHF database was inclusive of long-term ECG recordings (about 20 hours) from 15 CHF patients (11 men, 22-71 years old, and 4 women, 54-63 years old, New York Heart Association (NYHA) class 3-4). The NSRID includes 54 long-term ECG recordings (approximate 24 hours) of subjects in NSR, composed by 30 men aged 28.5 to 76, and 24 women aged 58 to 73 [4]. The ethic committee of the University of Electronic Science and Technology of China (UESTC) approved this study. 15 CHF patients and 15 healthy controls were used for analyses of the CHF detection method, in which the healthy group was randomly acquired from NSRID.

Based on those raw long-term ECG recordings, only 20-hour data of subjects in both CHF and NSR groups were obtained for the study. R-R intervals for each individual were calculated. Based on R-R interval data, the samples were checked and rejected if their values differ from the mean of the overall R-R interval level by one fold. After this preprocessing procedure, R-R interval time series for all subjects were segmented into eight different assessment lengths (representing the electrocardiographic monitoring duration), i.e., 15-hour, 7.5-hour, 4-hour, 2-hour, 1-hour, 30-minute, 10-minute, and 5-minute. Since a 50%-overlapping was taken between neighboring assessment lengths, the number of assessment lengths for each individual was 1, 4, 9, 19, 39, 79, 239, and 479, respectively.

III. METHODS

A. Feature Extractions

The clinically standard HRV measures in short-term observations (5-min) were utilized, followed by time domain features (SDNN, RMSSD, and pNN50), frequency domain features (VLF, LF, HF, and LF /HF), and Non-linear features (ApEn and SampEn) [7]. In addition to R-R interval features, both raw R-R interval time series, and mean and Standard Deviation (STD) of R-R intervals in certain assessment length were applied in this study.

B. Classifier Selections

The superior classifiers reported in previous studies (SVM and RF), and a new classifier in CHF classification studies, RNN, were developed.

The SVM classifier was introduced to train and test the established HRV model. A linear kernel function was adopted because it was proved to give an optimal classification accuracy (v.s. other nonlinear kernels) since it can reach more efficient computational time. The tube ε and regularization constant C in the ε-insensitivity loss function for minimizing the empirical error was set to 0.001 and 1, respectively.

In RF related works, similar with classifier configurations from Masetic and Subasi' studies, the RF algorithm included 100 trees, and each was constructed by a constant parameter that equalled to (the number of R-R samples in assessment length $\times$ the number of selected HRV features).

Long Short-term Memory (LSTM) was introduced into the standard RNN framework as the classifier for the comparative study of raw R-R intervals, mean and STD of R-R intervals, and HRV features [8]. In the LSTM unit, both *ReLu* and *tanh* functions were determined to activate nodes of input gate, cell activation vector, output gate, and forget gate. Three-layer RNN framework was proposed to pass the information of LSTM units from one layer to another. The weights of the hidden units were updated through the back propagation algorithm with the Adagrad optimizer. The RNN model computation and optimization were done using TensorFlow package. Other hyper parameters were constant throughout the experiment and defined as follows: learning rate = 0.001, a number of output gates = 2 and forget gate bias = 1.0.

C. Performance Evaluation

In order to avoid the data overlapping problem between training and testing data, the LOOCV was applied in order for the investigation of the CHF classification via short-time electrocardiographic monitoring (Figure 1).

Based on the data in an assessment length, the feature vectors of raw R-R intervals, mean and STD of R-R intervals, HRV measures in part, and HRV measures in full, were respectively calculated. The RNN classifier was chosen to comparatively analyze classification results from those constructed feature vectors. During the LOOCV, the numbers of assessment lengths for 29 subjects (the remaining 1 subject for the use of testing) were trial-by-trial feed to train the RNN model. The assessment lengths from the remaining 1 subject were applied in the testing session to generate an individual accuracy result, indicating that the CHF detection accuracy for such subject during certain electrocardiographic monitoring duration. The LOOCV process was iteratively repeated until all subjects are assessed for all different assessment lengths. Based on the individual accuracy results collected from all subjects, the overall performance, i.e., specificity, sensitivity, and accuracy, were analyzed for the classification performance of CHF detection methods with short-time electrocardiographic monitoring.

IV. RESULTS

First of all, the feature of mean and STD of R-R intervals is statistically analyzed. The One Sample t-Test result indicates that the mean and STD of R-R interval features of the CHF group for all assessment lengths (15-hour, 7.5-hour, 4-hour, 2-hour, 1-hour, 30-minute, 10-minute, and 5-minute) is significantly different from those of the NSR group ($p < 0.005$).

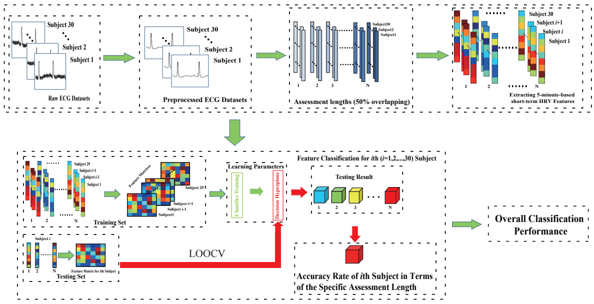

Fig. 1. LOOCV based performance evaluation procedure. The symbol i indicates the i^{th} subject from the datasets (totally 15 CHF patients and 15 NSR subjects), assessment lengths of whom will be tested in order to generate the accuracy rate for the i^{th} subject. Moreover, the averaged results from all subjects will be evaluated as the overall classification performance. The symbol N shows the number of trails of each subject for certain assessment length.

The statistical analysis was applied to short-term (5-min) HRV measures of both CHF and NSR subject groups. The One Sample t-Test outcome showed that only HRVs in both frequency-domain (VLF, LF, LF/HF) and nonlinear behaviours (SampEn) had significant differences ($p < 0.01$). In addition, RMSDD ($p = 0.059$) and ApEn ($p = 0.054$), had an edge to be significant. It was concluded that 6 HRV features indicated statistical differences occurred between two subject groups. Therefore, the HRV measures in part (HRVs only with statistically significant) and standard HRV measures in full (clinically standard HRV features) were then investigated in terms of the CHF detection via short-time electrocardiographic monitoring.

Table I illustrated the rates of accuracy, specificity, and sensitivity of all types of feature vectors combined with the multi-layer RNN classifier for all subjects. Introducing raw R-R intervals without feature extractions can achieve specificity of 100% and sensitivity of 86.67% associated with the 15-hour assessment length. This notable outcome revealed that long-term R-R intervals still preserve the valuable information to identify the CHF disease. However, accuracies of the model with raw RR intervals were dramatically dropped when the shorter assessment lengths were used. For instance, the sensitivity rate was dropped up to 66.67% for the 7.5-hour assessment length.

For mean and STD of R-R intervals, it has been noted that its classification performance are lower than those of raw R-R intervals, even in the 15-hour assessment length only acquiring 73.33% for specificity and 93.33% for sensitivity. More specifically, considering the 30-minute assessment length, the specificity was 57.57% (v.s. 65.40% for raw R-R intervals), and for NSR subjects with 80.95% (v.s. 75.34% for raw R-R intervals).

Comparing outcomes of HRV features in part to those of both raw R-R intervals and mean and STD of R-R intervals, HRV features in part had superior classification performance in both long-time (> 30 minutes) and short-time (≤ 30 minutes) assessment lengths. Sound classification performance with accuracy rates of 90% (15-hour), 88.89% (7.5-hour), 84.76% (4-hour), 87.78% (2-hour), and 87.31% (1-hour) could be obtained based on long-time electrocardiographic monitoring. Moreover, an accuracy rate of 87.09% at 30-minute for short-time electrocardiographic monitoring was still the best versus both of raw R-R intervals and mean and STD of R-R intervals.

Furthermoer, classification performance based on HRV features in full equally outperformed R-R interval-based features, given by the accuracy rate of 100% at 15-hour and 91.79% at 30-minute.

Following this work, comparative studies also investigated the classification performance of classifiers in recognizing CHF patients with different electrocardiographic monitoring durations. Outcomes of classifiers RNN, RF, and SVM were demonstrated while employing the standard HRV features in full as candidate feature vectors. The classification performance in terms of the variable assessment lengths were provided in Table II where the RNN classifier can receive the best outcomes in the long-time electrocardiographic monitoring (15-hour, 7.5-hour, 4-hour, and 2-hour) compared to those of RF and SVM. Furthermore, both classifiers of RNN and RF can offer acceptable outcomes in short-time electrocardiographic monitoring, showing the accuracy rates of 91.79% and 91.74% at 30-minute, respectively.

V. Discussions

The possibility of detecting the CHF disease via short-time electrocardiographic monitoring has not been well addressed in the previous literature. However, the quick CHF detection is extremely beneficial not only for medical experts who undoubtedly require fast reference advice in urgent medical conditions, but for any potential users who request for reliable and convenient alternatives for CHF early detections prior to the symptom occurrence.

Based on the survey of this study showed in Table I, it can be observed that the classification result in discriminating CHF patients varies from the assessment length, and utilizing HRV features could highly improve the classification performance and even reached the maximum accuracy of 100% when the ECG monitoring duration reached 15 hours. The best result for short-time electrocardiographic monitoring was yielded by using the feature vector of standard HRV features in full at the 30-minute assessment length with an accuracy rate of 91.79%. Results show that it was believed that a short-time electrocardiographic monitoring for the CHF detection could be feasible if the standard HRV features together with the classifier of RF or RNN are adopted. Based on these database, a specific 30-minute duration can be achieved by choosing HRV features in full.

The limitation of this study was the parameter selection for the RNN classifier. It should be improved in the future studies for the computational efficiency for both training and testing sessions, since in the real automated CHF diagnostic system time was the most important research concerns. Another limitation was the use of limited clinical CHF database. In the future studies, new CHF database with patients NYHA class I - V will be used for this study.

TABLE I

OVERALL CLASSIFICATION PERFORMANCE BY USING RAW R-R INTERVALS (NO FEATURE EXTRACTION), MEAN AND STD OF R-R INTERVALS, HRV FEATURES IN PART, AND HRV FEATURES IN FULL COMBINED WITH MULTI-LAYER RNN CLASSIFIER

Assessment Length	**RR intervals**			**Mean and STD of RR intervals**			**HRV features in part**			**HRV features in full**		
	ACC(%)	**SEN(%)**	**SPE(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**
15-hour	93.33	**100.00**	86.67	83.33	73.33	93.33	90.00	80.00	**100.00**	**100.00**	**100.00**	**100.00**
7.5-hour	76.67	66.67	86.67	76.67	64.44	88.89	88.89	82.22	95.56	**97.70**	**95.24**	**100.00**
4-hour	72.38	71.43	73.33	73.33	65.71	80.95	84.76	79.05	90.48	**95.57**	**95.92**	**95.24**
2-hour	70.67	66.22	75.11	71.55	64.89	78.22	87.78	82.67	92.89	**95.07**	**89.05**	**97.33**
1-hour	69.89	64.30	75.48	69.89	63.66	76.13	87.31	79.78	94.84	**91.43**	**87.33**	**95.27**
30-minute	70.37	65.40	75.34	69.26	57.57	80.95	87.09	81.27	92.91	**91.79**	**88.55**	**94.81**
10-minute	69.92	65.09	74.75	68.24	63.99	72.49	81.68	86.98	76.38	**89.52**	**87.18**	**91.71**
5-minute	69.91	64.16	75.66	68.29	64.50	72.08	82.74	74.56	90.93	**85.03**	**80.48**	**89.59**

TABLE II

CLASSIFICATION PERFORMANCE BY CLASSIFIERS OF MULTI-LAYER RNN, RF, AND SVM IN TERMS OF STANDARD HRV FEATURES IN FULL

Input Data	**9 HRV features of RR interval time series**								
Classifier	**Recurrent Neural Network**			**Random Forest Classifier**			**Support Vector Machine**		
Assessment Length	**ACC(%)**	**SEN(%)**	**SPE(%)**	**ACC(%)**	**SEN(%)**	**SPE(%)**	**ACC (%)**	**SEN(%)**	**SPE(%)**
15-hour	**100.00**	**100.00**	**100.00**	96.55	92.86	**100.00**	96.55	92.86	**100.0**
7.5-hour	**97.70**	**95.24**	**100.00**	96.55	92.86	100.00	94.25	88.10	**100.0**
4-hour	**95.57**	**95.92**	95.24	93.60	89.80	**97.14**	88.17	88.78	87.62
2-hour	**95.07**	89.05	**97.33**	94.02	**90.95**	96.89	87.59	87.62	87.56
1-hour	91.43	87.33	**95.27**	**91.55**	**89.17**	93.76	87.10	82.95	91.97
30-minute	**91.79**	88.55	94.81	91.74	**89.80**	93.54	85.55	75.28	**95.13**
10-minute	89.52	87.18	91.71	**90.90**	**89.43**	92.28	79.55	63.05	**93.02**
5-minute	85.03	80.48	89.59	**89.65**	**85.71**	**93.33**	82.76	78.57	86.67

VI. CONCLUSIONS

In this study, CHF detection method via short-time electrocardiographic monitoring was investigated. The databases from PhysioNet, BIDMC and NSRID, were introduced. For each individual, the R-R intervals were segmented into eight assessment lengths, i.e., 15-hour, 7.5-hour, 4-hour, 2-hour, 1-hour, 30-minute, 10-minute, and 5-minute. Based on the assessment length, raw R-R intervals, mean and STD of R-R intervals, HRV features in part, and HRV features in full were comparatively analyzed, while combining with classifiers, RF, SVM, RNN. The LOOCV was developed for performance verifications. Results showed that based on these database a specific 30-minute duration can be achieved by choosing HRV features in full (an accuracy rate of 91.79%). It was believed that a short-time electrocardiographic monitoring for the CHF detection could be feasible if the standard HRV features together with the classifier of RF or RNN are adopted. This study indicates that the short-time electrocardiographic monitoring approach can be applied for an automated CHF diagnostic system.

ACKNOWLEDGMENT

This work is supported by the Fundamental Research Funds for the Central Universities, China (grant No. ZYGX2015J118), and the National Natural Science Foundation of China (grant No. 51675087, 61522105).

REFERENCES

[1] L. Cardinale, A. M. Priola, F. Moretti, and G. Volpicelli, Effectiveness of chest radiography, lung ultrasound and thoracic computed tomography in the diagnosis of congestive heart failure, World journal of radiology,vol. 6, no. 6, p. 230, 2014.

[2] J. J. McMurray, S. Adamopoulos, S. D. Anker, A. Auricchio, M. Bohm, K. Dickstein, V. Falk, G. Filippatos, C. Fonseca, M. A. G. Sanchez et al., Esc guidelines for the diagnosis and treatment of acute and chronic heart failure 2012, Revista Portuguesa de Cardiologia (English Edition), vol. 7, no. 32, pp. 641642, 2013.

[3] J. P. Saul, Y. Arai, R. D. Berger, L. S. Lilly, W. S. Colucci, and R. J. Cohen, Assessment of autonomic regulation in chronic congestive heart failure by heart rate spectral analysis, American Journal of Cardiology, vol. 61, no. 15, pp. 12929, 1988.

[4] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation, vol. 101, no. 23, p. E215, 2000.

[5] G. Liu, L. Wang, Q. Wang, G. Zhou, Y. Wang, and Q. Jiang, A new approach to detect congestive heart failure using short-term heart rate variability measures. Plos One, vol. 9, no. 4, p. e93399, 2014.

[6] Z. Masetic and A. Subasi, Congestive heart failure detection using random forest classifier, Computer Methods & Programs in Biomedicine, vol. 130, no. C, p. 54, 2016.

[7] A. Hossen and B. Al-Ghunaimi, Identification of patients with congestive heart failure by recognition of sub-bands spectral patterns, Proceedings of World Academy of Science Engineering & Technolog, no. 44, p. 21, 2008.

[8] F. A. Gers and N. N. Schraudolph, Learning precise timing with lstm recurrent networks. JMLR.org, 2003.