A Deep Learning Approach to Monitoring and Detecting Atrial Fibrillation using Wearable Technology

Supreeth Prajwal Shashikumar[1*], Amit J. Shah[3,4,5], Qiao Li[2,3], Gari D. Clifford[2,3,6], Shamim Nemati[2,3]

Abstract— **Atrial Fibrillation (AF) is the most common cardiac arrhythmia in clinical practice, with a prevalence of 2% in the community. Not only it is associated with reduced quality of life, but also increased risk of stroke and myocardial infarction. Unfortunately, many cases of AF are clinically silent and undiagnosed, but long-term monitoring is difficult. Nonetheless, efforts at monitoring at-risk individuals and detecting clinically silent AF may yield significant public health benefit, as individuals with new-onset, asymptomatic AF would receive preventive therapies with anticoagulants and beta-blockers, for example.**

Wearables have enormous potential to provide low-risk and low-cost long-term monitoring of AF, but signals from such devices suffer from significant movement related noise that resembles AF. This work presents a robust approach to AF detection in a sequence of short windows with significant movement artifact. Pulsatile photoplethysmographic (PPG) data and triaxial accelerometry from 98 subjects (45 with AF and 53 with other rhythms) were captured using a multichannel wrist-worn device. A single channel electrocardiogram (ECG) was recorded (for rhythm verification only) simultaneously.

A novel deep neural network approach to classify AF from wrist-worn PPG signals was developed on this data. A continuous wavelet transform was applied to the PPG data and a convolutional neural network (CNN) was trained on the derived spectrograms to detect AF. Combining the output of the CNN with features calculated based on beat-to-beat variability and signal quality provided a significant accuracy boost.

Leave-one-out cross validation resulted in a pooled AUC of 0.95 (Accuracy=91.8%). The proposed approach resulted in a novel robust and accurate algorithm for detection of AF from PPG data, which is scalable and likely to improve in accuracy as the dataset size continues to expand.

I. INTRODUCTION

Approximately 1.5% - 2% [1] of the general adult population in the developed world has AF, an arrythmia which stems from irregular atrial contraction. Prevalence of AF in the general population is expected to increase over the coming years due to an aging population [2] [3]. AF is associated with a 5-fold increased risk of ischemic stroke, 3-fold increased risk of heart failure, and 2-fold increased risk of heart disease related death [4] [5] [6]. AF is a silent killer, as at least one-third of the patients are asymptomatic. As a result, stroke, heart failure, and death can be the first

presenting symptom of someone who has asymptomatic AF. This underscores the need for AF surveillance methods, which would allow early diagnosis and treatment. Because of the potentially large numbers of patients who may benefit from continuous AF screening, an affordable, continuous and portable monitoring system is needed.

Although ECG based devices have been the popular choice for AF detection until now [7] [8] [9], PPG based AF detection algorithms have also gained traction in the recent years. Owing to the increased popularity of wearable technologies, developing AF screening algorithms that run on the wearable devices is gaining attention. Since photoplethysmogram (PPG) is collected on most of these devices, AF detection algorithms based on PPG may potentially serve as a viable alternative to ECG based algorithms [10] [11]. Since wearable devices can be worn continuously with minimal burden, PPG-based AF detection algorithms are potentially advantageous over ECG-based methods, including patches and implantable devices, which are the current standard of care but are more short-term, costly, and sometime invasive.
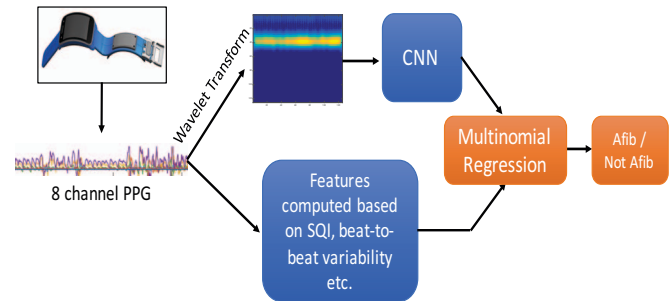

Fig. 1: An illustration of the AF detection algorithm. A 8-channel PPG is recorded using Samsung's Simband smartwatch [12]. The wavelet power spectrum obtained from 30 seconds window is fed into the CNN, the output of which is combined with the remaining features that are computed based on beat-to-beat variations and SQI to detect AF.

Although the PPG based algorithms (for wearable devices) show a lot of potential, we need to consider the difficulties associated with usage of such devices; first, PPG recordings from these devices can often be noisy due to the continuous movements of the users. Secondly, the occurrence of AF events can be rare and the noisy recordings can often mask such events. Thus, there is a need to develop algorithms that account for the aforementioned factors for AF detection.

In the recent years, Deep Neural Networks (DNN) have been used successfully in the field of Biomedical Signal pro-

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332

[2]Dept of Biomedical Informatics, Emory University, Atlanta, GA 30322.

[3]Dept of Medicine, Emory University School of Medicine, Atlanta, GA 30322.

[4]Dept of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322.

[5]Division of Cardiology, Veterans Affairs Hospital, Atlanta, GA 30322.

[6]Dept of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

[*]Corresponding author, Email: `supreeth@gatech.edu`
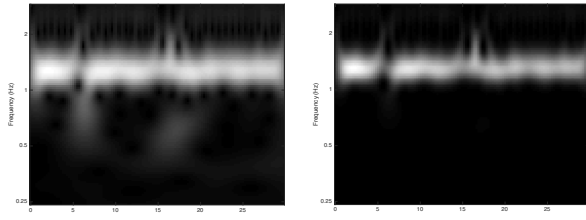
(a) Power spectrum without thresholding

(b) Power spectrum with thresholding

Fig. 2: Wavelet power spectrum for a 30s window of a subject with sinus rhythm

(a) Power spectrum without thresholding

(b) Power spectrum with thresholding

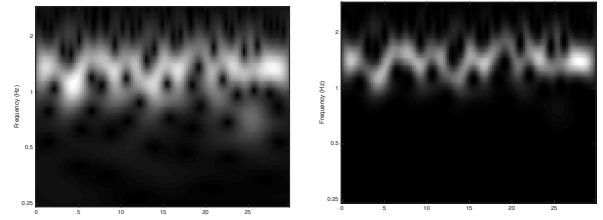Fig. 3: Wavelet power spectrum for a 30s window of a subject with AF

cessing. Research in the area of Biomedical Signal processing involves the study of signals like ECG, electroencephalogram (EEG), and PPG to predict wide range of physiological events in the human body. Some of the applications of DNNs in Biomedical Signal Processing include emotion recognition [13], seizure detection [14], and sleep stage detection [15]. In applications where imaging [16] or Spectral analysis is involved, CNNs have been employed successfully. In our work, we have harnessed the potential of CNNs to extract relevant information from spectral images obtained from PPG to detect AF.

In this paper, we extract spectral features from PPG (recorded using Samsung's Simband, a wearable wristband [12]) via wavelet analysis and then employ a CNN as a feature extraction tool to project the two dimensional wavelet spectrum to a one dimensional feature vector. Figure 1 illustrates the overall workflow of our AF detection algorithm. Our goal was to assess whether the feature vector obtained from the CNN, in combination with features based on beat-to-beat variability and Signal quality metrics, can achieve improved performance compared to the previous PPG based or ECG based approaches that rely solely on beat detection.

II. METHODS

A. Data Collection

All subjects were adult patients (18-89 years old) who were hospitalized and undergoing telemetry monitoring at the Emory University Hospital (EUH), Emory University Hospital Midtown (EUHM), and Grady Memorial Hospital (GMH). The study was approved by the institutional review board (IRB) of the Emory University Hospital and Grady Memorial Hospital in Atlanta, GA. Patients were recruited at random with an over-sampling of patients in AF; the rhythms were reviewed by an ECG technician, physician study coordinator, and cardiologist. The study took place from October 2015 through March 2016. Using a modern research watch-based wearable device (the Samsung Simband [12]), we recorded ambulatory pulsatile data from 98 subjects, 45 with AF and 53 with other rhythms for approximately 5 minutes. Single channel electrocardiogram (ECG), multi-wavelength photoplethysmography (PPG, 8 channels) and tri-axial acceleromtery (x, y, z) were recorded simultaneously at 128 Hz.

B. Preprocessing

Outlier rejection and amplitude normalization was performed on non-overlapping 30 seconds windows of each PPG channel. After subtracting the signal mean, we calculated the lower 5-percentile and upper 95-percentile of the PPG signal within each window segment, any value surpassing these two thresholds were set to the corresponding threshold (extreme value clipping), and the resulting segment was normalized by the maximum value in the segment. Next, each PPG channel was bandpass filtered to remove frequencies outside the range of $0.2 - 10$Hz, using an FIR filter of order 41. Within each window we performed PPG pulse onset detection using the slope sum function (SSF) approach [17]. Additionally, we calculated a signal quality index (SQI) for each window using the *Hjorth's purity* quality metric (SQI_{purity}) [18], which takes a value of zero for random noise and one for a sinusoidal signal. A second SQI was calculated by taking the average value of the magnitude of the accelerometer ($ACC = \sqrt{x^2 + y^2 + z^2}$) within each 30 seconds window (SQI_{mean_acc}). A third SQI was calculated by taking the standard deviation of ACC within each 30 seconds window (SQI_{STD_acc}).

C. Features based on PPG time series

Apart from the spectral information of PPG, we incorporated several metrics based on beat-to-beat interval variations in our AF detection algorithm. We calculated the sample entropy with the embedding dimensions $m = 1$, and 2 ($SampEn_1$ and $SampEn_2$, respectively). Next, we calculated the standard deviation of the beat-to-beat interval time series (STD), as well as a robust version of the standard deviation (STD_r), after discarding the intervals outside the 0.05-0.95 percentile range (we assumed that the extreme intervals are due to erroneous pulse onset detection). Finally, we calculated a weighted standard deviation feature, by taking the inverse of the ACC waveform within each interval as the weighing factor when calculating the standard deviation (STD_{acc}). Within each 30 seconds window, we used the SQI_{purity} to choose a channel with the highest signal quality and the features from this channel were used to represent the corresponding 30 seconds window. Next, we took the median of the features across all the 30 seconds windows as our final features (8 features). Additionally, we took the min and the max of the sample entropy features (4 additional features).

D. Wavelet Analysis

Wavelet transform is a spectral analysis technique in which a signal can be expressed as a linear combination of shifted and dilated versions of a "mother wavelet". By transforming the time series to time-frequency space, we can observe the variability in power of different frequencies across time. AF is a very common type of arrhythmia wherein the rhythm of the heart beat is irregularly irregular. Therefore, in the case of normal sinus rhythm the wavelet transform of the PPG will have dominant frequencies in the range of 1Hz - 2Hz [19] whereas in the case of AF, there will no more be a dominant frequency due to the irregular rhythm of the heart beat.

As mentioned in Section II-C, we found the channel with best signal quality (SQI_{purity}) for each of the windows. Next, we chose a window with the highest signal quality and the least ACC, after which the wavelet transform was applied on this 30 seconds window. The size of the wavelet spectrum computed was 125x125 and the mother wavelet used was the "Morlet wavelet". It can be observed from figure 2a and figure 3a that considerable amount of noise is present in the PPG recording which might corrupt the spectral content present in the PPG. Hence we performed further processing on the derived wavelet power spectrum in order to remove the contribution of noise to the spectrum. First, the statistical significance level of the wavelet power was estimated using a Monte Carlo method. Next, a large ensemble of surrogate data (N=100) was generated with the same first order autoregressive (R1) coefficients as the input signals. For each surrogate data we calculated the wavelet power, and used the 95-percentile of the power as the threshold above which the observed signal power was considered statistically significant (at 5% significance level). The mathematics behind the wavelet analysis is well documented in the references [20]. The wavelet power spectrum obtained after performing the thresholding was used as input to the CNN. Figures 2b and 3b show the wavelet power spectrum after performing the thresholding. Any significant power present in regions below 0.5Hz or above 2Hz might be due to the noise/artefacts present in the PPG signal, and the thresholding procedure clearly removes all such artefacts from the wavelet power spectrum.

E. Feature extraction using CNN

Since wavelet power spectrum sits in a 2 dimensional space, we used CNN to extract relevant information from the power spectrum and projected it to a smaller 1 dimensional feature vector. The input to the CNN was the thresholded wavelet power spectrum computed from a 30 seconds window of PPG. All the images were normalized by the max across all the spectrum images in the training set. Our CNN architecture, comprised of three successive convolutional layers and a max pooling layer, two pairs of two successive convolutional layer and a max pooling layer, and one fully connected layer. Each convolutional layer in the architecture was followed by a layer of activation with

TABLE I: List of Experiments performed and the corresponding AUC, Accuracy obtained

Experiment	AUC	Accuracy (in %)
Features alone	0.90	87.8
Wavelet power spectrum alone	0.89	87.7
Thresholded wavelet power spectrum alone	**0.92**	85.8
Wavelet power spectrum + Features	0.93	89.8
Thresholded wavelet spectrum + Features	**0.95**	**91.8**

Rectified Linear Unit (ReLU) nonlinearity. The CNN was implemented using the MatConvNet toolkit in Matlab [21].

All the convolutional layers in the CNN had filter size of 4x4 with a stride of 1 along both the directions, and all the pooling layers had pooling region of size 2x2 with a stride of 2 along both the directions. The number of filters used for each of the convolutional layers was 10, with the fully connected layer having a total of 50 filters. Thus, the output of the CNN was a 50 dimensional feature vector that represents the entire wavelet power spectrum.

F. Atrial Fibrillation Detection

Given the features derived in Section II-C combined with the feature vector obtained from the CNN, we built a *Elastic Net* logistic model to classify each patient into one of two classes of AF versus non-AF. The weights of the entire model were optimized using the RMSProp algorithm with "L1-L2" regularization imposed on the weights. The regularization parameter for the weights was chosen to be 0.001. The hyper-parameters of the entire model, including the regularization factors and the number of filters in the CNN, were optimized using Bayesian Optimization. The predicted probabilities of outcome for each test set example was pooled across all folds to calculate the final AUC.

III. RESULTS

In all of the experiments described below, we utilized a Leave One Out cross-validation study to assess the generalization performance of the AF detection algorithm ($N = 98$ patients). The list of all experiments performed and the corresponding performance of our AF detection algorithm has been tabulated in Table I

First, in order to understand the importance of spectral analysis in AF detection, we trained a model considering only the wavelet power spectrum and without using any of the features computed in Section II-C. Thus, the 50 dimensional vector obtained from the CNN was directly fed into the *Elastic Net* logistic model to perform AF classification. The pooled AUC obtained when the wavelet spectrum were fed into the CNN was 0.89 and the accuracy ($\frac{TP+TN}{TP+FP+FN+TN}$) was 87.7%. (Here TP is the number of true positive, TN is the number of true negative, FP is the number of false positive and FN is the number of false negative.) Whereas, the pooled AUC in the case when the thresholded wavelet spectrum are fed into the CNN was 0.92 and the accuracy was 85.8%.

Next, we used the features derived in Section II-C alone (the feature vector from CNN was excluded) to train an
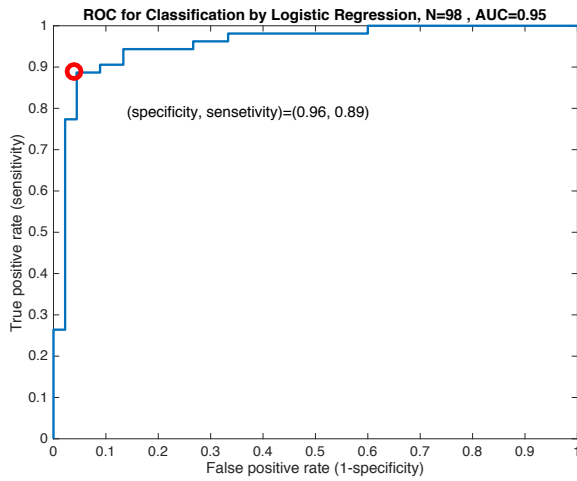
ROC for Classification by Logistic Regression, N=98 , AUC=0.95

Fig. 4: Receiver operating characteristic (ROC), based on Leave one out cross-validated performance on the holdout set

Elastic Net logistic model and we obtained a pooled AUC of 0.90 with an accuracy of 87.8%.

Finally, we combined the feature vector obtained from the CNN with features computed in Section II-C and fed into the *Elastic Net* logistic model. The pooled AUC obtained when the wavelet spectra were fed into the CNN was 0.93 and the accuracy was 89.8%. Whereas, the pooled AUC in the case when the thresholded wavelet spectra were fed into the CNN was 0.95 and the accuracy was 91.8%. Furthermore, in the case of thresholded wavelet spectrum we achieved a sensitivity ($\frac{TP}{TP+FN}$) of 0.89 for a specificity ($\frac{TN}{TN+FP}$) of 0.96. (See Figure 4) The accuracy corresponding to the above threshold was 91.8%.

IV. DISCUSSION AND FUTURE DIRECTION

In this work we presented a novel deep neural network approach to classify AF from wrist worn pulse data, using a CNN with a continuous wavelet transform of the PPG signal as the input. The proposed approach is not computationally very intensive due to the simple kernels being used in the CNN. We also note that combining the output of the CNN with the features calculated based on beat-to-beat variability and signal quality provided a significant accuracy boost. This is a key result - the naive application of a deep neural network to data was observed to be inferior to addition of carefully chosen features, particularly related to noise. In this way, the deep neural network is provided additional discriminatory power to separate out noise from AF data (which often appears similar to noise).

With increased data collection, we expect the promise of continued performance gains with the deep neural network architecture. However, an accuracy of 91.8% and and AUC of 0.95 is already comparable to ECG-based approaches on real (noisy) data - Oster and Clifford report state-of-the art AF detection accuracies of 90% on the Long Term AF database form PhysioNet, which rapidly degrade to 70-80% with the addition of typical amounts of noise observed in the ambulatory setting [22]. As such, this algorithm is highly promising for deployment in a real world setting.

REFERENCES

[1] A. J. Camm *et al.*, "2012 focused update of the esc guidelines for the management of atrial fibrillation," *European heart journal*, vol. 33, no. 21, pp. 2719–2747, 2012.

[2] Y. Miyasaka *et al.*, "Secular trends in incidence of atrial fibrillation in olmsted county, minnesota, 1980 to 2000, and implications on the projections for future prevalence," *Circulation*, vol. 114, no. 2, pp. 119–125, 2006.

[3] F. Rahman, G. F. Kwan, and E. J. Benjamin, "Global epidemiology of atrial fibrillation," *Nature Reviews Cardiology*, vol. 11, no. 11, pp. 639–654, 2014.

[4] S. Stewart, C. L. Hart, D. J. Hole, and J. J. McMurray, "A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the renfrew/paisley study," *The American journal of medicine*, vol. 113, no. 5, pp. 359–364, 2002.

[5] C. X. Wong, A. G. Brooks, D. P. Leong, K. C. Roberts-Thomson, and P. Sanders, "The increasing burden of atrial fibrillation compared with heart failure and myocardial infarction: a 15-year study of all hospitalizations in australia," *Archives of Internal Medicine*, vol. 172, no. 9, pp. 739–741, 2012.

[6] L. Friberg, M. Rosenqvist, A. Lindgren, A. Terént, B. Norrving, and K. Asplund, "High prevalence of atrial fibrillation among patients with ischemic stroke," *Stroke*, vol. 45, no. 9, pp. 2599–2605, 2014.

[7] F. Hobbs, D. Fitzmaurice, J. Mant, E. Murray, S. Jowett, S. Bryan, J. Raftery, M. Davies, and G. Lip, "A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. the safe study," *Health Technology Assessment*, vol. 9, no. 40, p. 93pp, 2005.

[8] E. Svennberg, J. Engdahl, F. Al-Khalili, L. Friberg, V. Frykman, and M. Rosenqvist, "Mass screening for untreated atrial fibrillation the strokestop study," *Circulation*, vol. 131, no. 25, pp. 2176–2184, 2015.

[9] M. Aronsson *et al.*, "Cost-effectiveness of mass screening for untreated atrial fibrillation using intermittent ecg recording," *Europace*, p. euv083, 2015.

[10] A. Carpenter and A. Frontera, "Smart-watches: a potential challenger to the implantable loop recorder?" *Europace*, p. euv427, 2016.

[11] S. Nemati *et al.*, "Monitoring and detecting atrial fibrillation using wearable technology," in *Engineering in Medicine and Biology Society (EMBC), 2016 Annual International Conference of the IEEE*. IEEE, 2016.

[12] "Samsung Simband's official website." [Online]. Available: https://www.simband.io/

[13] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[14] J. Turner, A. Page, T. Mohsenin, and T. Oates, "Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection," in *2014 AAAI Spring Symposium Series*, 2014.

[15] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Advances in Artificial Neural Systems*, vol. 2012, p. 5, 2012.

[16] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuznieecky, "Classification of patterns of eeg synchronization for seizure prediction," *Clinical neurophysiology*, vol. 120, no. 11, pp. 1927–1940, 2009.

[17] A. V. Deshmane, "False arrhythmia alarm suppression using ecg, abp, and photoplethysmogram," Ph.D. dissertation, Citeseer, 2009.

[18] S. Nemati, A. Malhotra, and G. D. Clifford, "Data fusion for improved respiration rate estimation," *EURASIP journal on advances in signal processing*, vol. 2010, p. 10, 2010.

[19] M. H. Beers, A. J. Fletcher, T. Jones, R. Porter, M. Berkwitz, and J. Kaplan, *The Merck manual of medical information*. Pocket Books, 2003.

[20] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear processes in geophysics*, vol. 11, no. 5/6, pp. 561–566, 2004.

[21] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for matlab," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.

[22] J. Oster and G. D. Clifford, "Impact of the presence of noise on RR interval-based atrial fibrillation detection," *Journal of Electrocardiology*, vol. 48, no. 6, pp. 947–951, 2015.