Recurrent Neural Network Circuit for Automated Detection of Atrial Fibrillation from Raw ECG

Sudarsan Sadasivuni[1], Rahul Chowdhury[1], Vinay Elkoori Ghantala Karnam[1],
Imon Banerjee[2*], and Arindam Sanyal[1*]
[1]Electrical Engineering Department, University at Buffalo, Buffalo, NY 14260, USA.
[2]Department of BioMedical Informatics, and Department of Radiology, Emory University, Atlanta, GA 30322, USA.
[*](co-senior authors); Email: ssadasiv@buffalo.edu

2021 IEEE International Symposium on Circuits and Systems (ISCAS) | 978-1-7281-9201-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISCAS51556.2021.9401666

***Abstract*—A recurrent neural network (RNN) is presented in this work for automatic detection of atrial fibrillation from raw ECG signals without any hand-crafted feature extraction. We designed a stacked long-short term memory (LSTM) network - a special RNN with capability of learning long-term temporal dependencies in the ECG signal. The RNN is digitally synthesized in 65nm CMOS process, and consumes 21.8nJ/inference at 1kHz operating frequency, while achieving state-of-the-art classification accuracy of 85.7% and f1-score of 0.82. The energy consumption of the proposed RNN is $8\times$ lower than state-of-the-art integrated circuits for arrhythmia detection.**

***Index Terms*—recurrent neural network, long-short term memory, electro-cardiograph, atrial fibrillation, health monitoring**

I. Introduction

It is estimated that 12.1 million people in the USA will suffer from atrial fibrillation (AFib) by 2030 [1]. AFib is the most common type of heart arrhythmia, and causes about 1 in 7 strokes [2]. While strokes caused by AFib tend to be more severe than strokes due to other underlying issues, many people suffering from AFib are asymptomatic which leads to reduced awareness and less chances of managing stroke risks for these patients. A potential solution for managing AFib risks is through continuous health surveillance using wearable devices that can monitor wearer electro-cardiograph (ECG) signal and identify AFib in real-time. Accuracy of AFib detection can be significantly improved by using artificial intelligence (AI) algorithms to analyze continuous ECG data in real time collected from wearables. However, commercial wearables rely on cloud-based AI models, and requires continuous transmission of sensor data to the cloud which is energy expensive, and also susceptible to unwanted exposure of user data during transmission [3]. Hence, a better solution is to perform AI analysis in the wearable itself. While several solutions have been proposed for AI circuits that detect arrhythmia from ECG signals [4]–[6], these techniques perform AI analysis on the traditional handcrafted features from raw ECG signal which can be energy expensive to compute. Furthermore, such features do not guarantee to be a good descriptor for the classification task.

The contribution of this work is to demonstrate a recurrent neural network (RNN) that detects AFib from raw ECG signal segments without hand-crafted feature extraction, and achieves state-of-the-art classification performance. Directly analyzing raw ECG signal instead of extracted features avoids information loss introduced during feature extraction by correlating the targeted classification task and saves the time and effort for manual feature engineering. Not performing feature extraction also reduces energy consumption of the proposed classifier. Fig. 1 shows an example application of the proposed RNN which will embedded inside an ECG sensor to detect AFib in real-time from segments of the ECG signal acquired by the sensor.

Stacked long-short term memory (LSTM) architecture is used for classification of ECG signal. Stacked LSTM is a RNN architecture which is capable of learning long-term dependencies in a time-series due to long-short memory and simultaneously generates reduced dimensionality via abstraction in stacked layers. Thus, such model is suitable for identifying arrhythmic patterns in the ECG signal over time. The proposed RNN is demonstrated on the 2017 PhysioNet public dataset [7] that comprises single lead ECG data from 7,068 patients. The RNN model is digitally synthesized in 65nm CMOS process, and consumes 21.8nJ/inference which is $8\times$ lower than state-of-the-art, while achieving better classification accuracy. The rest of this paper is organized as follows: Section II describes the dataset used in this work, Section III discusses design of the RNN, and comparison with state-of-the-art, while the conclusion is brought up in Section IV.

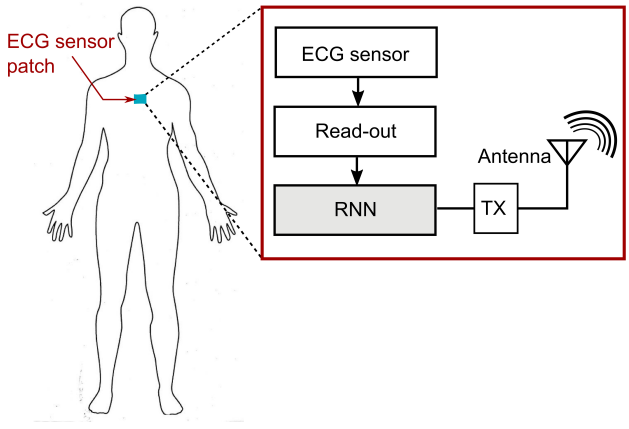

Fig. 1: Example application of the proposed work – embedding RNN inside ECG sensor for continuous and real-time AFib detection

978-1-7281-9201-7/21/$31.00 ©2021 IEEE

II. Description of the Dataset

The 2017 PhysioNet dataset comprises of ECG recordings lasting from 9 seconds to over 60 seconds and are collected using AliveCore device. The ECG recordings are sampled at 300Hz, and band-pass filtered by the AliveCore device. The ECG recordings contains normal or sinus rhythm, AFib and noisy data. Fig. 2 shows example ECG segments for normal, AFib and noisy classes.

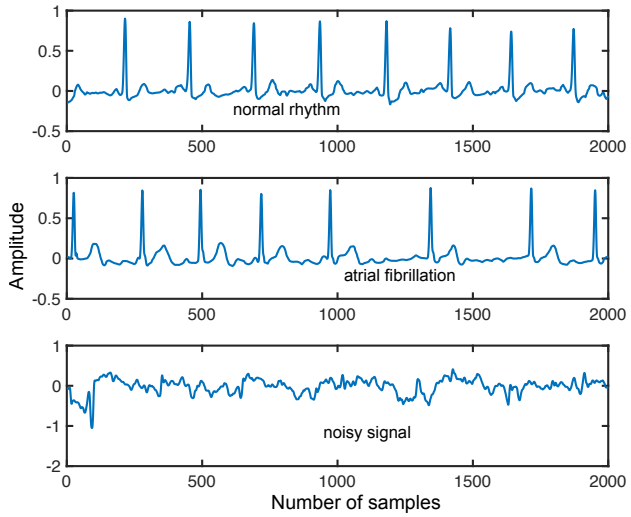

Fig. 2: Example ECG segments for normal, AFib and noisy classes

III. Proposed Recurrent Neural Network

A. Stacked LSTM architecture

A stacked LSTM is capable of learning long-term dependencies in temporal data due to long-short memory and simultaneously generates reduced dimensionality via abstraction in stacked layers. A single LSTM block comprises of 4 layers - a cell, an input gate, an output gate and a forget gate. The cell remembers temporal values and the three gates control the flow of information into and out of cell. Fig. 3(a) shows the schematic of LSTM block used in this work. The forget gate layer looks at the input X_t, output of prior LSTM unit, h_{t-1}, and passes their sum through a sigmoid function which decides how the cell state of the prior unit, C_{t-1} will be used for the current unit. The prior cell state is completely forgotten if the sigmoid output is '0' and fully retained if the sigmoid output is '1'. The input gate updates the current cell state C_t based on the output of forget gate and X_t and h_{t-1}. Finally, the output gate filters the current cell state using sigmoid and tanh functions, and produces the output, h_t.

Stacked LSTM is an extension to the traditional LSTM model that has multiple hidden LSTM layers where each layer contains multiple memory cells. By stacking the LSTM hidden layers we increase the depth of the network which ultimately helps to recombine the learned representation from prior layers and create new representations at high levels of abstraction. Fig. 3(b) shows the proposed RNN that utilizes stacked LSTM layers with inter-mediate drop-out layers and a fully connected layer followed by softmax layer for classification output. The fully connected layer performs weighted summation of the outputs of the second LSTM layer before passing the result to the softmax layer. In order to show the benefit of stacking the LSTM layers, we also compare the LSTM models with and without stacking in Section III-B.

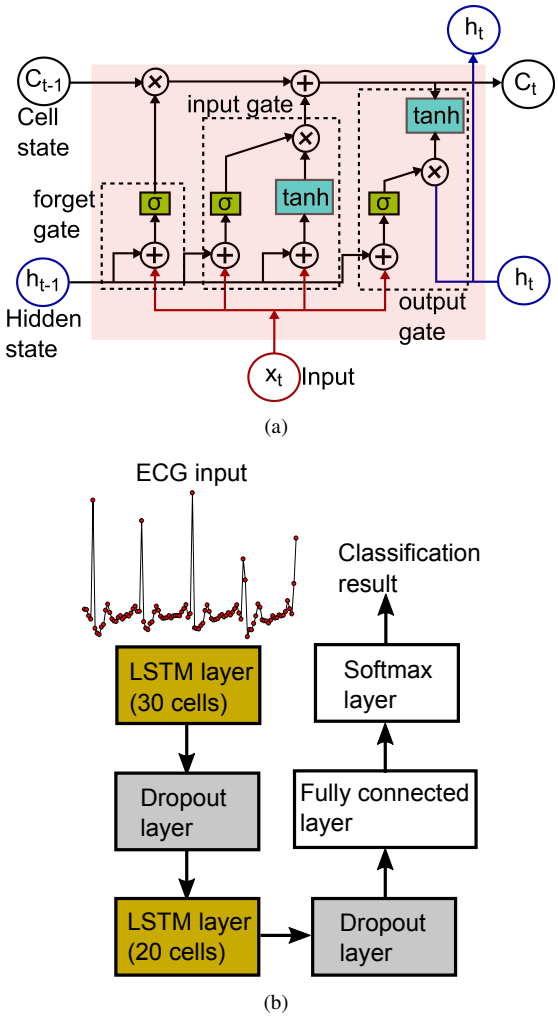

Fig. 3: (a) architecture of LSTM cell (b) architecture of RNN used in this work

B. Hyper-parameter tuning

The dataset is randomly partitioned into train, validation and test set, with 70% of the dataset used for training, 10% for validation and 20% is held-out for testing. Hyper-parameter tuning is performed by optimizing classification accuracy of the trained model on the validation set for different values of the hyper-parameters. The 2017 PhysioNet dataset has different length of ECG segments for each patient. The ECG segment length is swept to find the optimum segment length for classification, and the validation accuracy is plotted versus segment length in Fig. 4. For each segment length, zero-padding is used for ECG signals with less samples than the

segment length. The highest validation accuracy is obtained for ECG segment length of 9000.

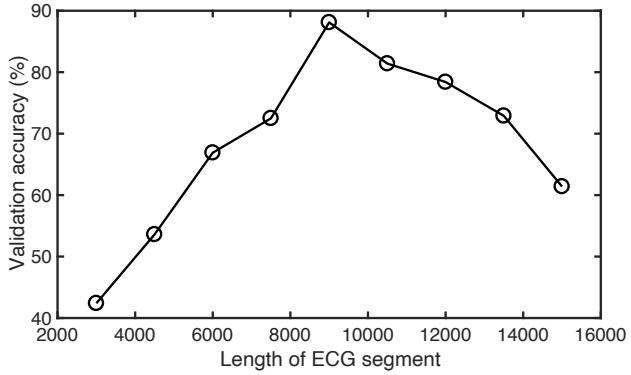

Fig. 4: Validation accuracy versus length of ECG segment

Other hyper-parameters of the RNN architecture - number of LSTM layers, number of cells in each LSTM layer and dropout percentage - are also fine-tuned to optimize validation accuracy. For an RNN with single LSTM layer, the validation accuracy is 51%, while the validation accuracy is 85% for an RNN with three stacked LSTM layers. The highest validation accuracy of 88.1% is obtained for an RNN with two LSTM layers. Fig. 5(a) shows the classification accuracy for different number of cells in the two LSTM layers. The highest validation accuracy is obtained for 30 cells in the first LSTM layer and 20 cells in the second LSTM layer. Fig. 5(b) shows validation accuracy versus dropout percentage in each dropout layer. Dropout is used to prevent over-fitting of the model. 30% dropout results in the highest validation accuracy.

C. Circuit design

The RNN model is synthesized digitally in 65nm CMOS technology using 32-bit fixed-point implementation. Fig. 6 shows layout of the synthesized RNN as well as the gate-level synthesized diagram of a single LSTM cell. The RNN occupies an area of 7.3mm^2 with 87% of the area occupied by memory cells associated with LSTM cells.

D. Results and comparison

Fig. 7 shows the simulated confusion matrix calculated on the test set. The classification accuracy of the RNN is 85.7%. Table I reports the precision, recall and f1-score [8] of the classifier on the 3 classes. The macro f1-score, precision and recall values are 0.83, 0.85 and 0.82 respectively.

TABLE I: Classification performance on test set

Class	Precision	Recall	f1-score
Normal	0.89	0.91	0.90
AFib	0.77	0.70	0.73
Noisy	0.83	0.82	0.83

Fig. 8 shows the classification accuracy versus bit width in the fixed-point synthesized RNN. The highest classification accuracy is obtained for 32-bit fixed point implementation.

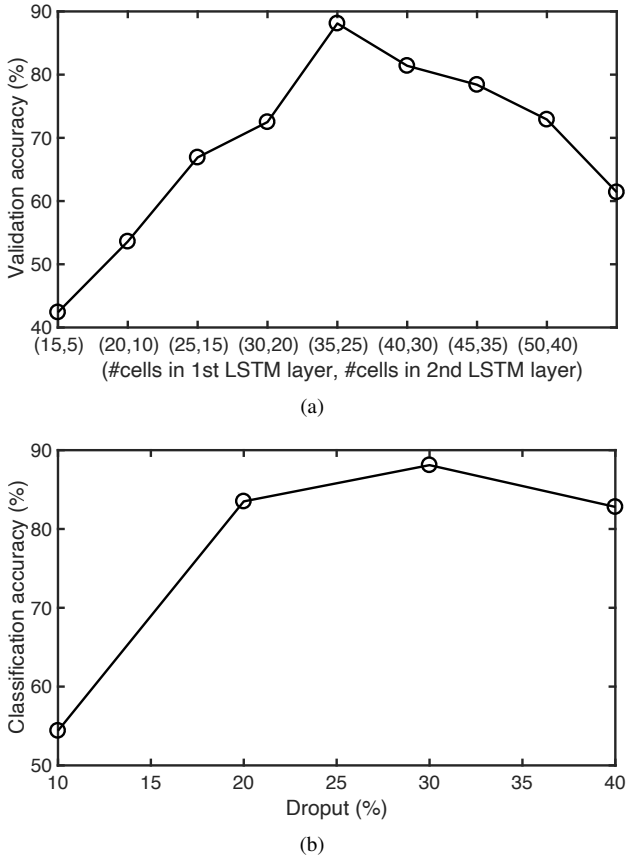

Fig. 5: Validation accuracy as a function of (a) number of cells in the LSTM layers (b) dropout percentage

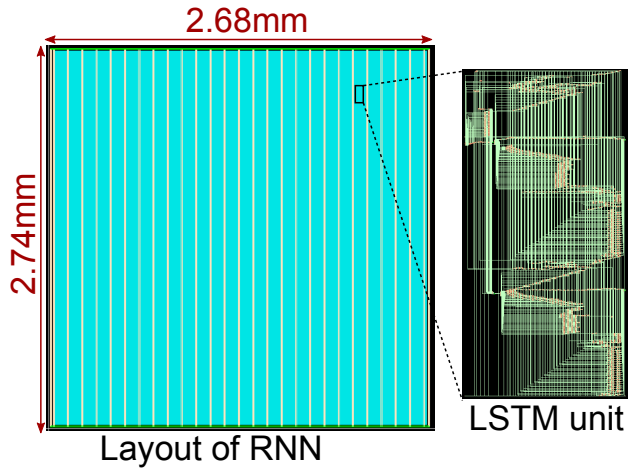

Fig. 6: Layout of the RNN with synthesized LSTM cell as inset

The RNN consumes 21.8nJ/inference and 2.9nJ/inference for 32-bit implementation and 16-bit implementation respectively while operating at 1kHz frequency from 1.2V supply.

Table II compares this work with state-of-the-art AI models on the 2017 PhysioNet dataset. The proposed RNN achieves state-of-the-art classification performance and f1-score without

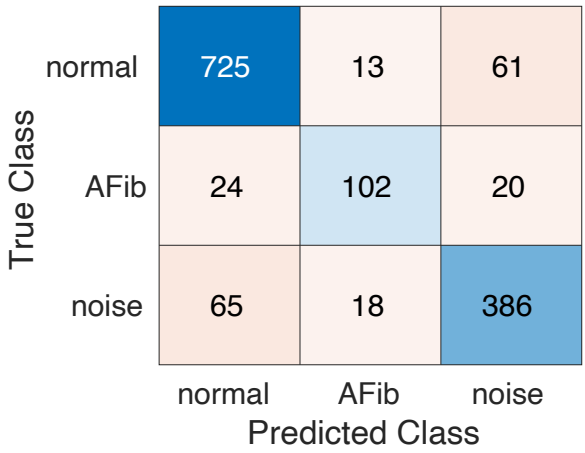

Fig. 7: Simulated confusion matrix on the test set

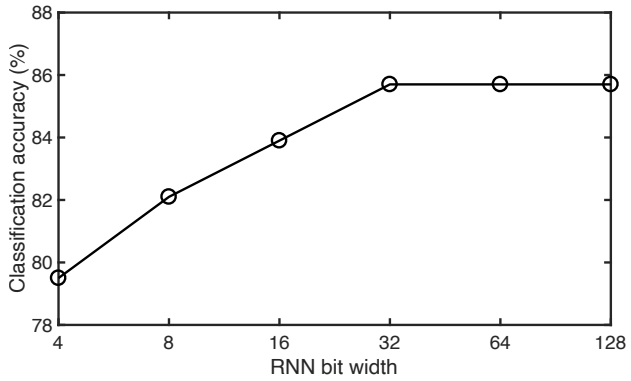

Fig. 8: Classification accuracy vs bit width

using feature extraction in contrast to the existing models. High performance of the proposed work is likely due to the fact that feature extraction introduces some information loss through reduction in data dimensions. In contrast, the raw ECG signals are directly used for classification in our work which allows the RNN model to use all the information in the ECG signal. Removing feature extraction block also reduces power and area of our RNN and makes it suitable for integration on sensor node.

TABLE II: Comparison with AI models

	Model	Features #	f1-score
[9]	ANN	169	0.79
[10]	SVM	42	0.80
[11]	SVM	50	0.81
[12]	Decision tree (DT)	30	0.82
[13]	Ensemble DT	37	0.75
[14]	CNN	–	0.71
This work	RNN	–	0.82

Table III compares the performance of our synthesized RNN with state-of-the-art arrhythmia detection circuits which has similar classification complexity as AFib detection. The state-of-the-art arrhythmia detectors all perform feature extraction on ECG signal before classification. The proposed RNN achieves $8\times$ and $64\times$ lower energy than state-of-the-art for 32-bit and 16-bit implementations respectively, thanks to the direct processing of raw ECG signals using stacked LSTM.

TABLE III: Comparison with state-of-the-art ASICs

	Model	Energy	Process
JSSC'13 [4]	SVM	124μJ[1]	130nm
JSSC'13 [5]	SVM	186nJ[1]	90nm
JSSC'19 [6]	MLP	330nJ[1]	65nm
This work	RNN (32b)	21.8nJ[2]	65nm
	RNN (16b)	2.9nJ[2]	65nm

[1]:Measured results; [2]post-layout simulated results

IV. Conclusion

This work has presented a stacked LSTM based RNN for detecting AFib from ECG signal directly. The proposed digital RNN consumes only 21.8nJ/inference which is an enabling factor for integration of the RNN into an ECG sensor for real-time continuous monitoring of user health. Since the proposed RNN is completely digital, it is robust against noise and environmental variations, and its energy consumption can be further reduced through technology scaling.

Acknowledgment

This material is based on research sponsored by Air Force Research Laboratory under agreement number FA8650-18-2-5402. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

References

[1] 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," *The American journal of cardiology*, vol. 112, no. 8, pp. 1142–1147, 2013.

[2] P. A. Wolf, R. D. Abbott, and W. B. Kannel, "Atrial fibrillation as an independent risk factor for stroke: the Framingham Study." *Stroke*, vol. 22, no. 8, pp. 983–988, 1991.

[3] J. Liu and W. Sun, "Smart attacks against intelligent wearables in people-centric internet of things," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 44–49, 2016.

[4] K. H. Lee and N. Verma, "A low-power processor with configurable embedded machine-learning accelerators for high-order and adaptive analysis of medical-sensor signals," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 48, no. 7, pp. 1625–1637, 2013.

[5] S.-Y. Hsu *et al.*, "A 48.6-to-105.2 μW machine learning assisted cardiac sensor SoC for mobile healthcare applications," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 49, no. 4, pp. 801–811, 2014.

[6] S. Yin, M. Kim, D. Kadetotad, Y. Liu, C. Bae, S. J. Kim, Y. Cao, and J.-s. Seo, "A 1.06-μ W Smart ECG Processor in 65-nm CMOS for Real-Time Biometric Authentication and Personal Cardiac Monitoring," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 8, pp. 2316–2326, 2019.

[7] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, "AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017," in *IEEE Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[8] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1, no. 10.

[9] F. Andreotti, O. Carr, M. A. Pimentel, A. Mahdi, and M. De Vos, "Comparing Feature-Based Classifiers and Convolutional Neural Networks to Detect Arrhythmia from Short Segments of ECG," in *IEEE Computing in Cardiology (CinC)*, 2017.

[10] J. A. Behar, A. A. Rosenberg, Y. Yaniv, and J. Oster, "Rhythm and Quality Classification from Short ECGs Recorded Using a Mobile Device," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[11] L. Billeci, F. Chiarugi, M. Costi, D. Lombardi, and M. Varanini, "Detection of AF and other rhythms using RR variability and ECG spectral measures," in *IEEE Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[12] G. Bin, M. Shao, G. Bin, J. Huang, D. Zheng, and S. Wu, "Detection of atrial fibrillation using decision tree ensemble," in *IEEE Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[13] P. Bonizzi, K. Driessens, and J. Karel, "Detection of Atrial Fibrillation Episodes from Short Single Lead Recordings by Means of Ensemble Learning," in *IEEE Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[14] B. Chandra, C. S. Sastry, S. Jana, and S. Patidar, "Atrial fibrillation detection using convolutional neural networks," in *Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.