






In-Sensor Artificial Intelligence and Fusion With Electronic Medical Records for At-Home Monitoring

Sudarsan Sadasivuni , Member, IEEE, Monjoy Saha , Sumukh Prashant Bhanushali , Imon Banerjee ,
and Arindam Sanyal , Member, IEEE

Abstract—This work presents an artificial intelligence (AI) framework for real-time, personalized sepsis prediction four hours before onset through fusion of electrocardiogram (ECG) and patient electronic medical record. An on-chip classifier combines analog reservoir-computer and artificial neural network to perform prediction without front-end data converter or feature extraction which reduces energy by $13\times$ compared to digital baseline at normalized power efficiency of 528 TOPS/W, and reduces energy by $159\times$ compared to RF transmission of all digitized ECG samples. The proposed AI framework predicts sepsis onset with 89.9% and 92.9% accuracy on patient data from Emory University Hospital and MIMIC-III respectively. The proposed framework is non-invasive and does not require lab tests which makes it suitable for at-home monitoring.

Index Terms—Sepsis, artificial intelligence, in-memory computing, data fusion, artificial neural network, reservoir-computer.

I. INTRODUCTION

SEPSIS is a life-threatening medical condition that arises when the body initiates an extreme response to an infection in the bloodstream. Real-time, at-home monitoring of at-risk patients using smart wearable is a potential solution for predicting sepsis onset and timely intervention. This work presents an artificial intelligence (AI) framework that combines patient electronic medical record (EMR) and electrocardiogram (ECG) data to automate risk prediction of sepsis onset without requiring a clinical expert in the loop. ECG is selected as the

Manuscript received 12 June 2022; revised 27 October 2022 and 10 January 2023; accepted 19 February 2023. Date of publication 2 March 2023; date of current version 19 May 2023. This work was supported in part by the National Science Foundation under Grant CCF-1948331 and in part by Air Force Research Laboratory under Grant FA8650-18-2-5402. This paper was recommended by Associate Editor Y. Zheng. (Corresponding author: Sudarsan Sadasivuni.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Emory University Institutional Review Board under Application No. MOD005-IRB00033069, and performed in line with the Helsinki Declaration.

Sudarsan Sadasivuni is with the Department of Electrical Engineering, University at Buffalo, Buffalo, NY 14260 USA (e-mail: ssadasiv@buffalo.edu).

Monjoy Saha is with the National Institutes of Health, MD 20814 USA (e-mail: monjoybme@gmail.com).

Sumukh Prashant Bhanushali and Arindam Sanyal are with the School of Electrical, Computer, Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: spbhanus@asu.edu; arindam.sanyal@asu.edu).

Imon Banerjee is with the Mayo Clinic, Phoenix, AZ 85054 USA (e-mail: banerjee.imon@mayo.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TBCAS.2023.3251310>.

Digital Object Identifier 10.1109/TBCAS.2023.3251310

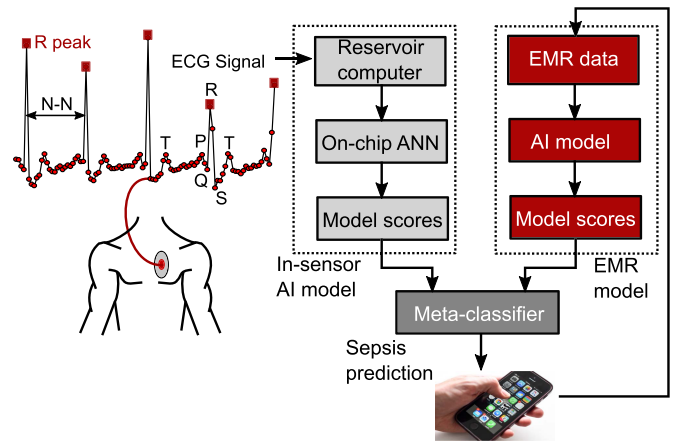


Fig. 1. Overview of the proposed AI framework for fusion of sensor and EMR data for sepsis onset prediction.

physiological modality in this work since studies [1], [2], [3], [4] have shown that sepsis patients show prolonged duration and reduced amplitude for QRS segments which can be picked up by AI models. A recent work [5] has also shown that deep-learning models can reliably perform sepsis screening using only patient ECG signal. The proposed framework is shown in Fig. 1, and comprises three components – a) in-sensor processing AI circuit for analyzing ECG signal and predicting risk of sepsis onset; b) a classifier that predicts risk of sepsis onset from EMR – patient demographics (age, gender, race and ethnicity) and co-morbidity data; and c) a meta-learner that combines prediction results from ECG and EMR to predict risk of sepsis onset with high accuracy. Wireless transmission of continuous sensor data is energy inefficient since information rate of ECG signal is much lower than its sampling rate. While there are several techniques to compress RF transmission [6], [7], [8], [9], [10], [11], [12], the compression ratio is typically limited to $< 20\times$. In this work, we propose a AI-driven technique that uses local, in-sensor AI to classify ECG segments and transmit prediction score instead of raw data that can compress transmission data by a much higher factor than existing compression techniques and significantly reduce sensor energy. However, integrating computationally intensive AI classifier into a resource constrained sensor is challenging. The majority of attempts [13], [14], [15], [16], [17], [18], [19], [20] to reduce energy consumption of AI circuits use a) in-memory/near-memory computing b) reduced precision computations. To address this energy bottleneck in wireless

bio-medical sensors, we propose an analog signal processing neural network that directly processes analog ECG samples.

The key contributions of this work are - a) on-chip analog classifier comprising of a reservoir-computer (RC) followed by a 3-layer artificial neural network (ANN) that process analog ECG segments while reducing energy consumption by $13\times$ compared to digital baseline (front-end ADC followed by digital ANN) and reduces overall sensor energy by $159\times$ compared to direct transmission of digitized ECG segments; b) a fusion model that combines patient single-modality physiological signal and demographics to predict sepsis onset with high accuracy without requiring multiple modality sensor data and laboratory test results as in current state-of-the-art sepsis onset prediction works. The proposed contributions are demonstrated on sepsis dataset obtained from Emory University Hospital (EUH) and MIMIC-III dataset. The AI models are trained on EUH dataset and tested on both EUH and MIMIC-III dataset. With the approval of Emory Institutional Review Board (IRB), de-identified sepsis dataset is obtained from EUH. The cohort consisted of 800 patients admitted to the ICUs at two hospitals within the Emory Healthcare system in Atlanta, Georgia from 2014 to 2018. The dataset has 53.3% sepsis patients and 46.7% non-sepsis patients, with male/female split for sepsis and non-sepsis patients of 54.7%/45.3% and 47.4%/52.6% respectively. The EUH dataset is randomly partitioned into 720 training samples and 80 test samples. The AI models trained on EUH dataset is tested on publicly available MIMIC-III dataset which comprises of de-identified, comprehensive clinical data of patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. The MIMIC-III dataset has 4559 patients with 40.2% sepsis patients and 59.8% non-sepsis patients, male/female split for sepsis and non-sepsis patients of 53.8%/46.2% and 55.8%/44.2% respectively. In a prior work [21], we have demonstrated that fusion of prediction scores from ECG using in-sensor ANN and EMR achieves state-of-the-art prediction accuracy for detection of sepsis onset. The key differences of this work with our prior work [21] are - 1) this work presents an analog-to-information conversion technique that directly analyzes analog ECG samples without digitizing and feature extraction; 2) the results in [21] are based on circuit simulation of ANN whereas this work presents measurement results on test-chips fabricated in 65 nm CMOS; 3) the proposed in-sensor RC+ANN and fusion techniques are validated on two datasets. The AI models for analyzing EMR data and fusion of prediction scores from in-sensor RC+ANN and EMR model are implemented in software.

The rest of this paper is organized as follows: Section II presents the details of the in-sensor AI architecture and circuit design, and the EMR model, Section III presents measurements results with 65 nm prototypes on the EUH and MIMIC-III datasets and comparison with state-of-the-art, and Section IV brings up the conclusion.

II. ON-CHIP AI CLASSIFIER

A. Challenges With Prior Works and Motivation

RF transmission is the largest contributor of wireless sensor power, and hence, local in-sensor signal processing is

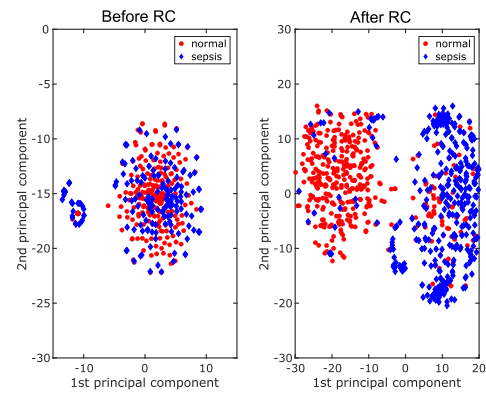


Fig. 2. t-SNE plot showing the distribution of ECG samples before passing through the reservoir and after passing through the reservoir layer.

preferred to continuous RF transmission [11]. As an example, a low-power MedRadio transmitter consumes $67\mu\text{W}$ power [22] which is significantly higher than on-chip feature extraction [11], [23]. While RF transmission has benefited from techniques that reduce transmission data rate, such as sparsity-based data compression algorithms [6], [7], [8], derivative-based adaptive sampling [9], level-crossing sampling [10], and adaptive resolution digitization [11], [12]. The aforementioned techniques have reported compression of transmission data by $2 \sim 16\times$. In contrast to the prior techniques, we propose to embed AI in the sensor itself to analyze each ECG segment and transmit only prediction score instead of ECG data or extracted features which results in significant reduction in transmission volume and reduce sensor energy by $159\times$. The transmitted prediction score is fused with prediction scores from patient EMR to produce personalized prediction for each patient.

Fig. 3(a) shows the block diagram of the in-sensor AI for analyzing ECG signals. Continuous-time ECG signal, $X(t)$, is sampled by the input layer and multiplied with input weight matrix to produce an output \vec{V} . \vec{V} is then sent to the reservoir-layer of the RC, and outputs of the reservoir layer, \vec{R}_k , is sent to an ANN for generating prediction score for the ECG segment under analysis.

B. Reservoir-Computer Design

RC is a well-known computing paradigm that uses static nonlinearity to project the input signal to high-dimensional space, thus allowing easier separation of different input classes. No training is performed in the input or reservoir layers, and the weights are drawn from random distribution. Fig. 2 shows an example of the separability of classes due to RC. Since the data is high-dimensional, t-distributed stochastic neighbor embedding (t-SNE) plot is used for visualization. t-SNE plot of the input ECG samples show that the sepsis and normal classes are clustered together. After processing of the ECG signals through RC, the two classes show good separation. While reservoir computing was invented almost two decades earlier and has been extensively used in the machine-learning literature, hardware implementation of reservoir computing have been mostly on optics/photonics platform with few analog silicon

implementations [24], [25], [26]. In contrast to prior silicon RC, the proposed RC is based on the architecture in [27] and does not require large capacitors to realize biological time-constants which is energy-inefficient, and does not require background calibration for analog delay elements or nonlinearity element.

Output of the RC with N reservoir neurons can be mathematically expressed as

$$\vec{R}_k[n] = H \left(G_i \vec{W} \times \vec{X}^T + G_f \vec{W}_r \times \vec{R}_k[n-1] \right) \quad (1)$$

where $\vec{X} = [X_1 X_2 \dots X_D]$ is analog ECG input with D samples, \vec{W} is $N \times D$ input weight matrix, \vec{W}_r ($D \gg N$), \vec{W}_r is $N \times N$ inter-connection weight matrix for the reservoir layer, $H(\cdot)$ is nonlinear activation for RC, G_i is input scaling factor and G_f is feedback gain. As in [27], identity matrix is used for \vec{W}_r which simplifies the hardware implementation since \vec{W}_r can be realized using a single-cycle delayed feedback. The restriction on \vec{W}_r is consistent with [28] which has shown through systematic investigation that a simple reservoir architecture with sparsely inter-connected reservoir provides comparable accuracy as more complicated reservoir architectures. G_i and G_f and N are set to 0.6, 0.1 and 63 respectively to optimize prediction accuracy and ensure stability of the reservoir, $D = 6000$ corresponding to 20 s ECG segments. The design parameters are selected based on hyperparameter optimization using 10-fold cross validation on training data.

Fig. 3(b) shows the circuit schematics of a single channel (j -th channel out of N channels) of the input and reservoir layers and the corresponding timing diagram. The input layer performs vector matrix multiplication of \vec{W} and D samples of the ECG signal over a time-period of $T_s/2$ (where T_s is the sampling period) which is split into D cycles of complementary clock phases ϕ_1 and ϕ_2 [29]. For the j -th channel, the input layer performs element-wise multiplication of entries in the j -th row of the input weight matrix \vec{W} with the corresponding ECG samples over D cycles, where the j -th row of \vec{W} is denoted as W_j . The input layer is realized using a switched-capacitor integrator which is reset every sampling period (T_s). The ECG signal is sampled at a speed of DF_s (where $F_s = 1/T_s$). During each ϕ_1 phase, the ECG signal is multiplied with an element in W_j and the result is sampled on the input capacitor C_{in} as shown in Fig. 3(b). Since elements of W_j are set to '0/1,' the multiplication of ECG signal and elements of W_j is implemented using 2 switches controlled by W_j and complement of W_j as shown in Fig. 3(b). Thus, for the J -th channel and k -th ϕ_1 phase ($k \in [1, D]$), and using a notation of $W[j, k]$ to denote the corresponding element of \vec{W} , if $W[j, k] = 1$, the ECG signal is sampled on C_{in} and if $W[j, k] = 0$, C_{in} is discharged. During the next ϕ_2 phase, the charge on C_{in} is transferred to the integration capacitor C_{int} . Thus, after D -cycles of ϕ_1 and ϕ_2 , the integration capacitor holds the product of $W_j \times \vec{X}^T$ scaled by C_{in}/C_{intg} which is the input scaling factor G_i . The voltage on the integration capacitor is sampled at ϕ_3 as V_j which is the input to the j -th reservoir neuron. The integration capacitor is then discharged after D cycles of ϕ_1 and D cycles of ϕ_2 . An operational-transconductance amplifier (OTA) is used to sum

input to the reservoir layer with delayed feedback from the reservoir neuron. Output of the OTA represents the term within parenthesis in (1) and is passed through the nonlinearity $H(\cdot)$ which is implemented using a feed-forward common-source amplifier as shown in Fig. 3(b). The non-linear activation function $H(\cdot)$ is based on Mackay-Glass nonlinearity. Output of the nonlinearity circuit is buffered and drives a 10-bit successive approximation register (SAR) ADC, and its delayed output is feedback to the input OTA through a resistive digital-to-analog converter (R-DAC). The reservoir layer is time-multiplexed to save on-chip area such that one physical neuron is used to realize N virtual neurons by operating the reservoir layer at NF_s where $F_s = 1/T_s$ is the frequency of operation of the RC+ANN and ECG input is sampled at DF_s . The ADC is used in the reservoir loop for accurate generation of N -cycle delay in the time-multiplexed feedback path since generation of precise analog delay is difficult in practice. The RC input layer is off-chip for this design to allow testing with different \vec{W} .

In contrast to conventional analog design, the circuits components in the RC can be nonlinear since all nonlinearity is absorbed into the reservoir dynamics. Relaxed linearity requirements allow amplifiers in the input layer and RC as well as the ADC to be low bandwidth, which results in increased nonlinearity due to slewing and incomplete settling, but reduces both noise and power. Fig. 4 shows the simulated accuracy from ECG analysis as a function of settling time in the switched-capacitor input layer, and time-constant of the amplifier (τ) is set to $T_s/8/D$. Size of the sampling capacitor in the input layer is set by noise and accuracy requirements. Input-referred noise in the input layer can be shown to be

$$\overline{V_{n,in}^2} = \frac{kT}{C_{in}} \frac{1}{1-\beta} + \frac{1}{\beta} \frac{kT}{C_{eq}} \left(\frac{4}{3} \right) \left(\frac{C_{intg}}{C_{in}} \right)^2 \quad (2)$$

where β is the feedback factor and $C_{eq} = C_L + (1-\beta)C_{intg}$, C_L is the load capacitor. Keeping the feedback capacitor C_{intg} fixed to 400fF and load capacitor of 100fF, the sampling capacitor C_{in} is swept and the RC+ANN accuracy and energy of the input layer are plotted in Fig. 5. Based on the simulation results, C_{in} is set to 10fF.

The lower bound on bandwidth of amplifiers in the reservoir layer is set by stability requirements. Since the reservoir is strongly nonlinear, the RC loop has to be linearized around its operating point to theoretically analyze stability. The worst-case scenario from stability perspective occurs when the RC loop has the highest gain, corresponding to the highest gain of the nonlinearity function $H(\cdot)$ that occurs for the smallest input seen by the nonlinearity circuit. The highest possible gain for $H(\cdot)$ is found through simulations for different values of feedback gain, G_f . Fig. 6(a) shows the discrete-time, linearized model of the RC with G_h denoting gain of $H(\cdot)$. The summing amplifier and the unity-gain buffer in Fig. 3(b) uses the same OTA with unity-gain bandwidth of ω_1 and feedback factor of the summing amplifier is β , and 3-dB bandwidth of the nonlinearity circuit is ω_2 . Stability of the RC is analyzed by finding the roots of (3)

$$1 + \frac{z^{-3}}{(1-k_1 z^{-1})(1-k_1 z^{-1})(1-k_1 z^{-1})} = 0 \quad (3)$$

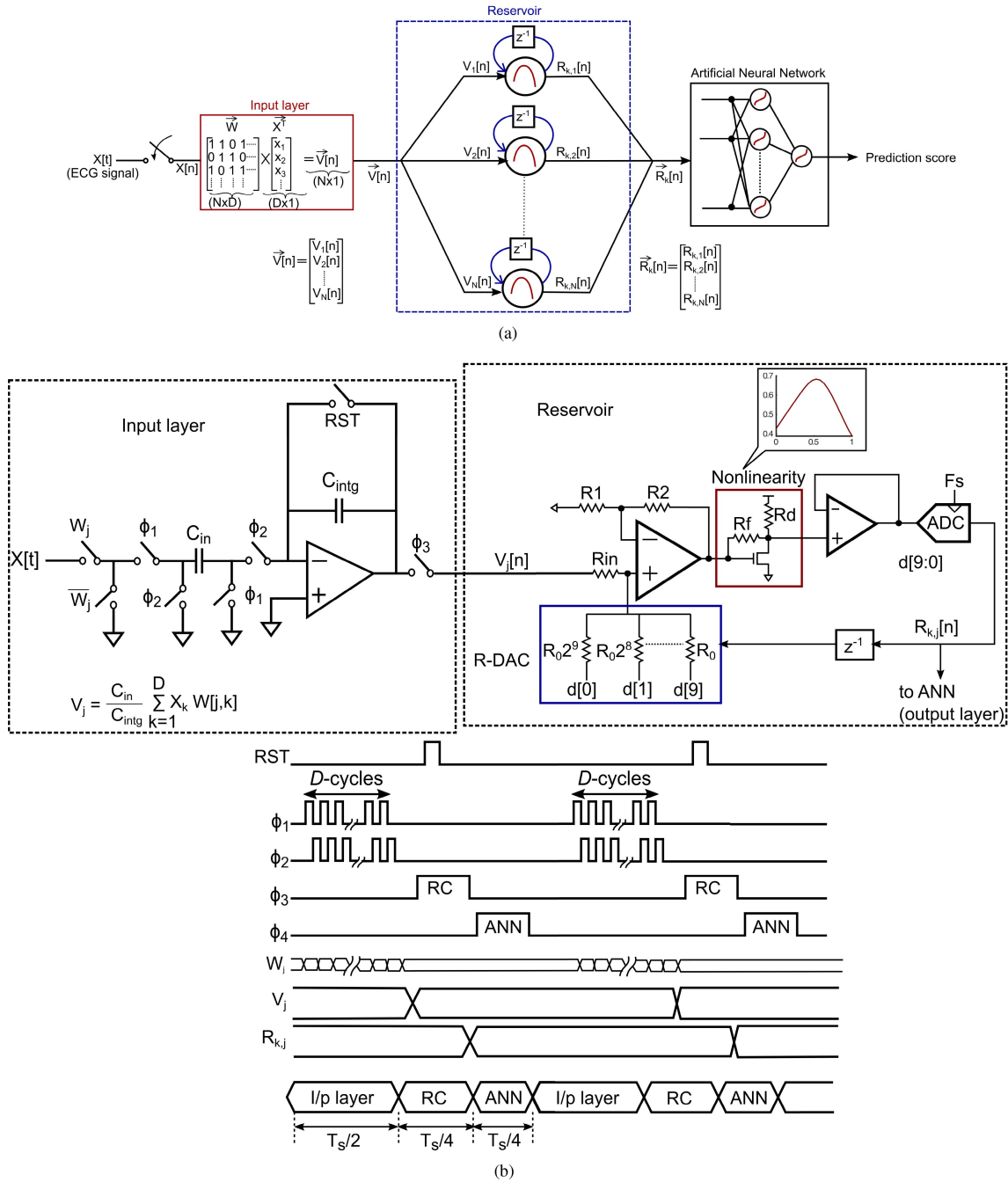


Fig. 3. a) Architecture of reservoir-computer for analyzing ECG signals b) circuit schematic of the input and reservoir layers.

Fig. 6(b) plots stability contours versus normalized values of ω_1 and ω_2 as a function of β . The stable region shrinks as G_f increases, and ω_1, ω_2 reduce. ω_1 and ω_2 are set to $2\pi \times 0.9F_s$ ($2\pi \times 0.9NF_s$ after time-multiplexing) for $G_f = 0.1$ to ensure a wide stability margin.

C. ANN Model Training and Circuit Design

The ANN has 20 neurons in the first hidden layer, and 6 neurons in the second hidden layer. The hidden layers use custom tanh activation function, while the output layer uses a custom softmax activation function. The voltage output of the softmax

function is compared with a threshold voltage (V_{th}) to generate the ANN decision, i.e., non-sepsis/sepsis. The activation circuits are designed using single-stage, common-source differential amplifiers as shown in Fig. 7. Multiply-and-add (MAC) operations in the hidden and output layers are performed using switched-capacitor circuits in which the ANN model weights are encoded as capacitance values. The switched-capacitor circuits operate in 2 non-overlapping and complementary phases - sampling (ϕ_s) and evaluation (ϕ_{sb}). During ϕ_s all the capacitors are discharged by connecting both terminals to the common-mode voltage V_{cm} . During the evaluation phase, the input activations, X_p and X_m , are applied to the top-plate of the capacitors.

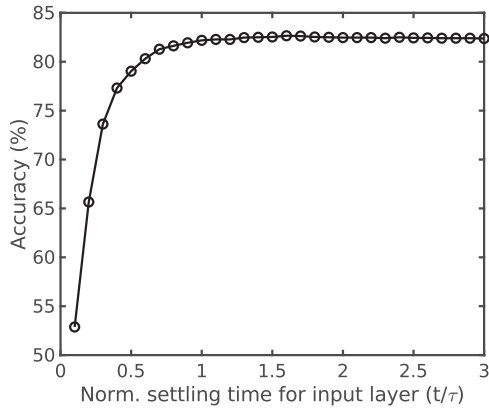


Fig. 4. Simulated accuracy vs input layer bandwidth .

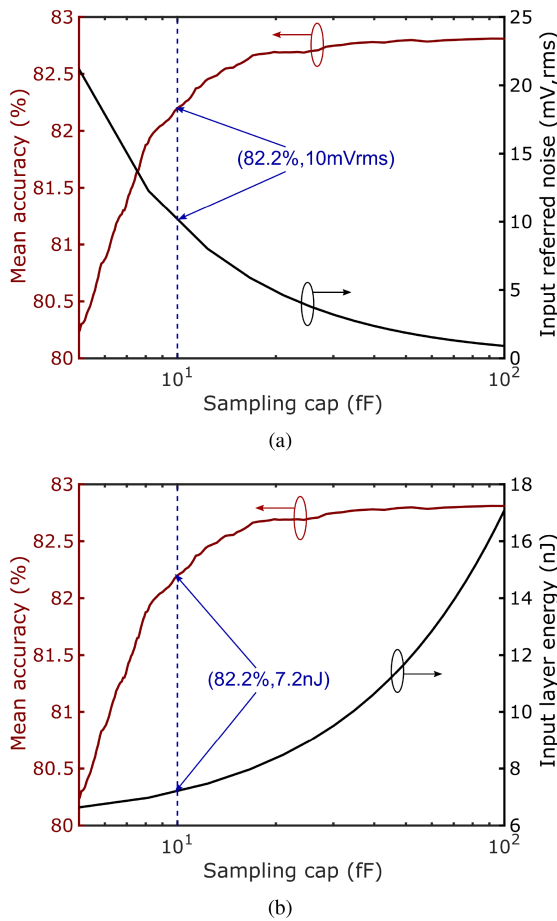


Fig. 5. Simulated a) accuracy and noise; and b) energy of input layer versus sampling capacitor.

The voltages at the bottom-plate of the capacitors at the end of evaluation phase are given by $(\sum X_p[k]C[k]) / \sum C[k]$ and $(\sum X_m[k]C[k]) / \sum C[k]$ respectively where $k \in [1, n]$ and n is the number input activations for the ANN layer. Thus, at the end of the evaluation phase, the differential input voltage to the amplifiers is the result of the MAC operations with input activations and network weights. The MAC result is then passed through the nonlinear activation function realized by

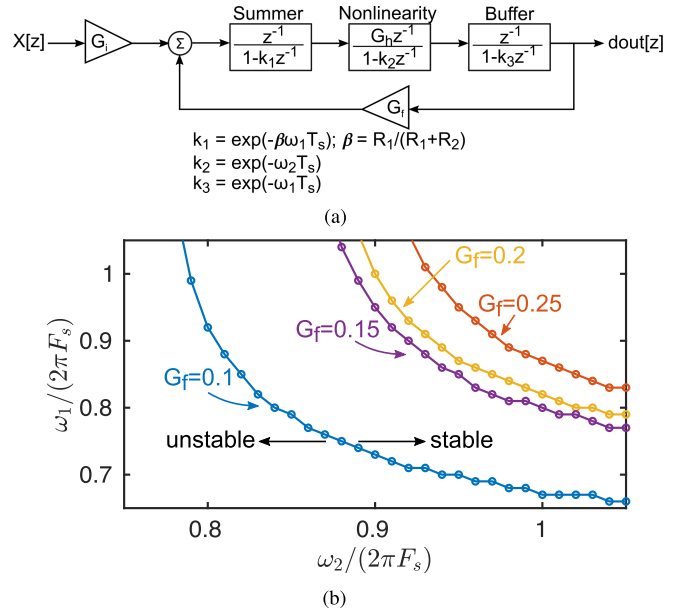


Fig. 6. a) Linearized model of the RC b) stability contours.

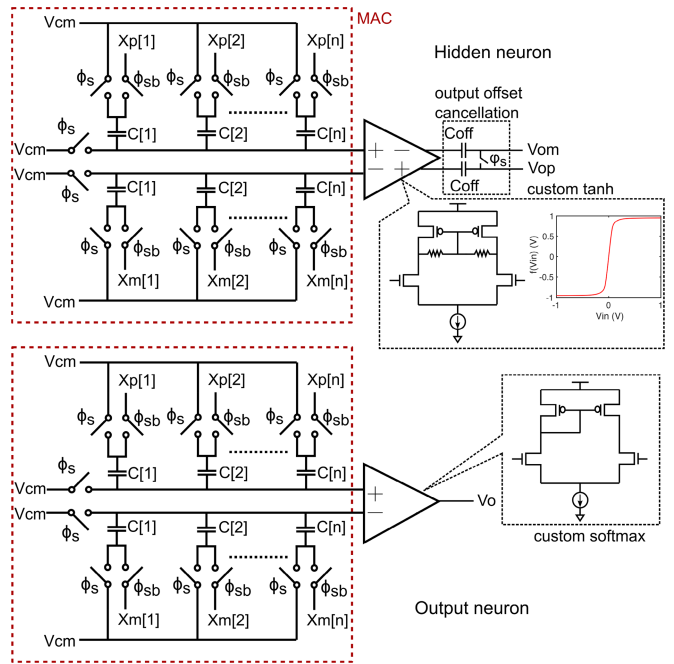


Fig. 7. Circuit schematic of custom hidden and output neurons.

the differential amplifier to produce the output activation. The fully differential amplifiers in the hidden layers use output offset cancellation technique to reduce amplifier offset. As shown in Fig. 7, the inputs to the differential amplifiers are shorted to V_{cm} during ϕ_s and the output offset is stored in the capacitors C_{off} . During the evaluation phase, the offset storage capacitors are in series with the amplifier outputs and remove the offset from the output activation. Offset in the output layer is removed through foreground calibration as described later. The custom analog activation functions resemble their ideal, mathematical counterparts, but are not exactly the same. To ensure good

matching between software ANN model and IC measurements, we use a hardware-software co-design methodology in which amplifier transfer curves, and their derivatives, are used to train the ANN model iteratively [30]. Stochastic gradient descent is used to optimize the ANN model by minimizing the loss function at each epoch. Once the ANN is fully trained, the model weights are encoded as capacitor values as shown in Fig. 7. The ANN weights are quantized to 4-b in the hidden layers, and 6-b in the output layer. The weight quantization is done during the training iterations to minimize effect of quantization error. A 4fF unit capacitor is used to realize an LSB weight without degrading ANN accuracy due to capacitor mismatch.

D. EMR Model

In addition to ECG data, we incorporated patient demographics and co-morbidities for prediction of sepsis onset using an EMR model. We performed a series of pre-processing steps on the EMR data prior to analysis and model development. As the first data pre-processing step, we applied standard data cleaning steps, including removing empty cells, special characters. Label encoding technique is used to convert categorical demographic features to numerical quantities. The co-morbidity data is in the form of ICD-10 codes and is converted to vector format using Term Frequency- Inverse Document Frequency (TF-IDF) algorithm [31] that computes a score for each word in proportion to its significance in the corpus. The ICD-10 codes does not include any diagnosis codes and only captures prior co-morbidities of the patient, such as presence of diabetes, chest pain, abdominal pain, shortness of breath, seizures, chronic kidney disease, elevated blood pressure etc. A word t in the j -th ICD10 code is represented in the following vector format

$$V(t,j) = TF(t,j) \times IDF(t) \quad (4)$$

where $TF(t,j)$ is the normalized term frequency and is given by

$$TF(t,j) = \frac{\text{Number of times term } t \text{ appears in } j}{\text{Total number of terms in } j} \quad (5)$$

and $IDF(t)$ is the inverse document frequency that is given by

$$IDF(t) = \log \left(\frac{N}{DF(t)} \right) \quad (6)$$

where $DF(t)$ is the number of documents containing t . The TF-IDF tokenizer is trained on the training dataset for the vectorization. Finally the numeric representation of the categorical features and TF-IDF representation of the co-morbidities are combined using linear concatenation, and normalized by removing the mean and scaling to unit variance. Given the static nature of EMR, a single-point prediction model is used and optimal value of the hyper-parameters is tuned through 10-fold cross validation on the training data. Accuracy of different models - linear support vector machine, random forest and logistic regression - on the test set are 49%, 76% and 53% respectively.

E. Fusion Model

Fig. 8 shows the fusion model that performs decision level late fusion by aggregating predictions from the RC+ANN and

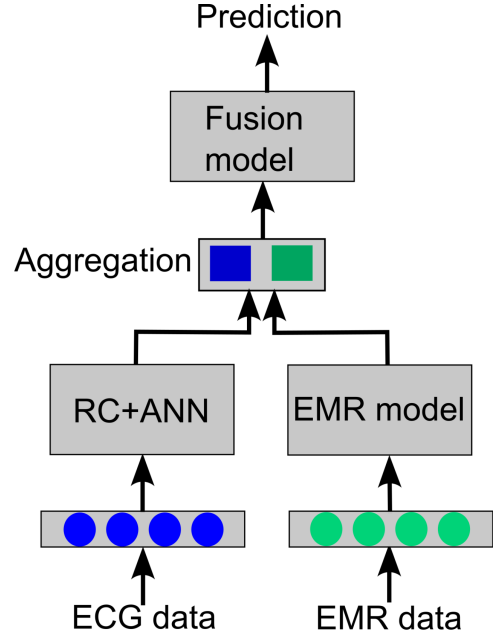


Fig. 8. Block diagram showing the fusion model that aggregates prediction scores from RC+ANN and EMR models.

EMR models. Linear support vector machine (SVM) is used as the meta-learner. Similar to the EMR model, hyperparameters of the meta-learner is tuned through 10-fold cross-validation on the training data. The linear SVM meta-learner is trained with square hinge loss function with L2 penalty and regularization parameter (λ) of 1. Mathematically, this can be described as solving the following optimization problem

$$\arg \min \frac{1}{2} \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^n \ell(y_i w^T x) \quad (7)$$

where x_i and y_i , ($i \in [1, n]$) are predictor and response variables of instances in the training dataset, w is the linear SVM classifier, and ℓ is the loss function defined by $\ell(\alpha) = \max\{0, (1 - \alpha)^2\}$.

III. MEASUREMENT RESULTS

Fig. 9 shows the measurement setup and energy breakdown. The RC and ANN chips are fabricated separately in 65 nm CMOS process, and integrated on printed circuit board level for lab measurement. National Instrument data acquisition (DAQ) module is used to load input data from a computer, and output of the ANN is captured using an oscilloscope and sent to the computer. A Matlab interface is used for communication between the computer, test chips and the NI DAQ. The ANN has a core area of 1.67 mm² and the RC has a core area of 0.24 mm². The on-chip reservoir layer consumes 2nJ/inference and the ANN consumes 7nJ/inference while the off-chip reservoir input matrix multiplier consumes 8.4nJ/inference [29] at 1.2 V supply and operating at $F_s = 1$ kHz. The energy for communication between the test chips is not included since this will be amortized once the two chips are integrated on the same die.

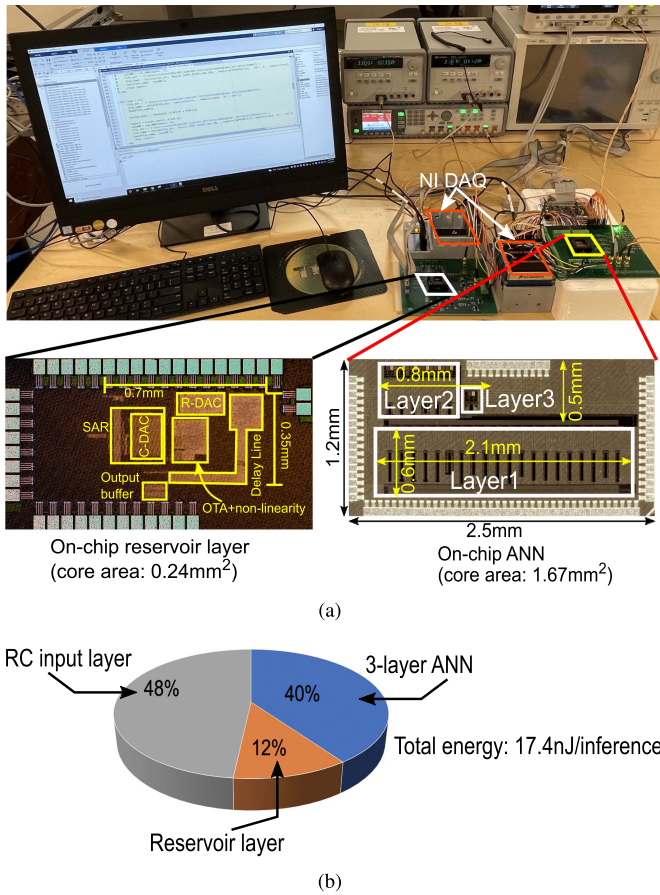


Fig. 9. a) Lab measurement setup with die microphotograph of the RC and ANN chips (b) energy breakdown.

A. Prediction Results With Emory and MIMIC-III Dataset

Fig. 10 shows the measured accuracy of the RC + ANN on the EUH test data. Offset in amplifier in the output layer of the ANN is calibrated by applying the training samples to the test-chip and setting the decision threshold voltage to maximize prediction accuracy on training samples. Fig. 10a) shows the measured accuracy on the hold-out test data for on-chip AI circuit with all chips calibrated and with chips 2-4 using the offset from chip 1. The proposed RC + ANN combine detects sepsis with mean accuracy of 78.9% and standard deviation of 0.5% from ECG signal 4 hours before onset if all chips are calibrated, while the mean accuracy drops slightly to 78.5% if only chip 1 is calibrated. The measurement results show small chip-to-chip variation and indicates robustness of the proposed RC+ANN test-chip. The mean accuracy improves to 89.9% after fusion with demographics and co-morbidity data. Linear SVM is used as meta-classifier for fusion. Sepsis onset prediction is more accurate closer to onset as shown in Fig. 10b). Table I shows the performance metrics (accuracy, specificity and sensitivity) for 4 test-chips as a function of time-to-onset after calibration, while Table II shows the performance metrics for the test-chips after fusion. Sensitivity is the measure of the AI model's ability to correctly diagnose patients with sepsis while specificity is the measure of the AI model's ability to correctly diagnose patients

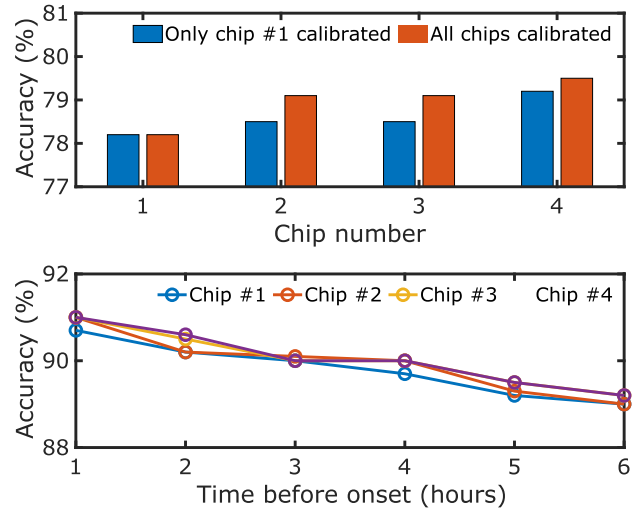


Fig. 10. a) Measured accuracy at 4 hours before onset with all chips calibrated, and with only chip 1 calibrated and all chips using the calibrated offset setting from chip 1 b) accuracy after fusion as a function of time before onset.

TABLE I
PERFORMANCE METRICS OF RC+ANN TEST-CHIPS

Chip	Time before sepsis onset					
	6 hr	5 hr	4 hr	3 hr	2 hr	1 hr
Chip 1						
Accuracy (%)	75.7	78.0	78.2	80.0	81.2	82.5
Sensitivity	0.66	0.71	0.66	0.74	0.72	0.72
Specificity	0.87	0.85	0.92	0.86	0.92	0.92
Chip 2						
Accuracy (%)	77.4	78.5	79.1	79.2	80.0	80.9
Sensitivity	0.68	0.66	0.70	0.71	0.73	0.72
Specificity	0.89	0.91	0.88	0.87	0.87	0.90
Chip 3						
Accuracy (%)	78.5	78.7	79.1	79.4	80.1	80.3
Sensitivity	0.68	0.69	0.69	0.69	0.71	0.73
Specificity	0.88	0.88	0.90	0.89	0.89	0.89
Chip 4						
Accuracy (%)	77.9	78.5	79.5	79.7	80.2	81.4
Sensitivity	0.68	0.68	0.68	0.70	0.72	0.73
Specificity	0.89	0.90	0.90	0.90	0.89	0.89

TABLE II
PERFORMANCE METRICS OF FUSION MODEL

Chip	Time before sepsis onset					
	6 hr	5 hr	4 hr	3 hr	2 hr	1 hr
Chip 1						
Accuracy (%)	89.0	89.2	89.7	90.0	90.2	90.7
Sensitivity	0.87	0.92	0.92	0.92	0.88	0.90
Specificity	0.90	0.87	0.88	0.87	0.92	0.91
Chip 2						
Accuracy (%)	89.0	89.3	90.0	90.1	90.2	91.0
Sensitivity	0.89	0.87	0.91	0.89	0.88	0.91
Specificity	0.89	0.92	0.89	0.90	0.92	0.91
Chip 3						
Accuracy (%)	89.2	89.5	90.0	90.0	90.5	91.0
Sensitivity	0.88	0.86	0.89	0.89	0.88	0.90
Specificity	0.91	0.92	0.90	0.91	0.92	0.92
Chip 4						
Accuracy (%)	89.2	89.5	90.0	90.0	90.6	91.0
Sensitivity	0.89	0.89	0.88	0.89	0.89	0.92
Specificity	0.89	0.89	0.92	0.91	0.91	0.91

TABLE III
COMPARISON WITH STATE-OF-THE-ART SOFTWARE AI MODELS FOR SEPSIS PREDICTION

	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	This work	
Data source	Physionet'19 challenge	Duke University Hospital	Singapore govt. hospital	Christina Care Health	MIMIC-III					EUH	MIMIC-III
Model	LSTM ¹	RNN ²	Random forest	LSTM+CNN	RNN	CNN	DL-ATT ³	Cox	LSTM	Fusion	
time-to-onset	4 hr	4 hr	4 hr	3 hr	7 hr	6 hr	4 hr	4 hr	1 hr	4 hr	
Accuracy	84.5%	—	—	91.5%	—	84.7%	—	64%	—	89.9% ⁴	92.9% ⁴
Sensitivity	—	0.84	0.87	0.97	0.88	0.87	0.49	0.89	0.85	0.9 ⁴	0.95 ⁴
Specificity	0.66	0.80	0.89	0.86	0.84	0.86	—	0.90	0.64	0.90 ⁴	0.89 ⁴
Vitals	8	9	5	6	10	0	7	10	6	1	
Lab tests	26	39	6	37	6	13	17	30	27	0	
EMR ⁵	6	36	4	35	0	0	3	19	3	5	

¹long-short term memory; ²recurrent neural network; ³attention-based deep-learning model; ⁴average of 4 test-chips; ⁵demographics and co-morbidities

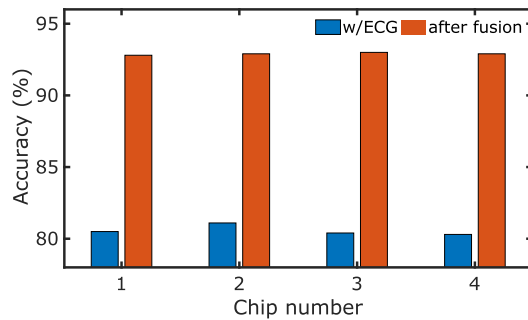


Fig. 11. Performance of 4 test chips before and after fusion.

that are normal. A highly sensitive AI model has low false negative results while a highly specific AI model has low false positive results. Mathematically, sensitivity is defined as the ratio of true positives to the sum of true positives and false negatives, while specificity is defined as the ratio of true negatives to the sum of true negatives and false positives. It is desirable to have both high sensitivity and high specificity.

All 4559 patient EMR and ECG data from MIMIC-III dataset is applied to the RC+ANN test-chips and the fusion AI models without re-training for predicting sepsis 4 hours before onset, and the results are summarized in Fig. 11. The test-chips achieve mean accuracy of 80.6% with ECG data and 92.9% after fusion. Standard deviation of accuracy of the 4 test-chips with ECG data is only 0.36% which indicates robustness of the proposed RC+ANN test-chips.

B. Comparison With State-of-The-Art and Discussion

Table III compares the proposed fusion model with state-of-the-art software AI models for different dataset. The proposed technique compares favorably with state-of-the-art using single modality sensor data source and no laboratory test results, which demonstrates feasibility of the proposed technique for at-home monitoring and is a key differentiation from state-of-the-art which requires multiple modality sensor data and/or laboratory test results. Table IV compares efficiency (TOPS/W) of the RC+ANN with state-of-the-art in-memory computing AI accelerator macros. The proposed RC+ANN achieves competitive power efficiency as state-of-the-art AI accelerators and matrix

multiplier macros even after including energy for data movement and output activations. Table IV also compares area efficiency of the AI accelerator macros in terms of area of a bitcell normalized to the CMOS technology in which the bitcell is designed. For this work, area of unit switched-capacitor circuit in the ANN (see Fig. 7) is considered which is equivalent of a bitcell in SRAM designs. The unit switched-capacitor in this design has larger area than capacitor-less SRAM bitcells, but has comparable area efficiency as bitcell designs that embed capacitor inside the bitcell. The trade-off with switched-capacitor based MAC operation is higher linearity than current-domain MAC [43] at the cost of area. It should be pointed out that a limitation of our implementation of RC+ANN is that the weights of the trained ANN cannot be changed after chip fabrication unlike SRAM based accelerators in which the neural network weights can be easily re-programmed. Fig. 12 compares the proposed RC+ANN with conventional technique of digitizing ECG segments and transmitting the digitized data, and digital baseline which performs feature extraction on digitized ECG segment followed by digital ANN before transmission of prediction scores. Transmission energy is assumed to be state-of-the-art 38 pJ/bit [44], and the ADC for digitizing ECG segment is assumed to consume 5fJ/conversion-step at 1 kHz and 12-bit resolution [45]. RC+ANN reduces energy/inference by 13× compared to digital baseline at 3% loss in accuracy, and by 159× compared to conventional technique. Fig. 12(b) plots energy/inference of recent state-of-the-art AI ICs for different bio-medical applications [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62]. The proposed RC+ANN technique consumes the lowest energy/inference which is 4× lower than state-of-the-art.

In this work, cleaned ECG signal from two dataset has been used to demonstrate the proposed techniques of in-sensor AI and fusion with EMR. However, in practical at-home monitoring applications, the acquired ECG signal is likely to contain artifacts and will required analog front-end (AFE) with band-pass filtering before the ECG signal is sent to the RC+ANN combination. The AFE will consume additional power that will reduce the energy advantage of the proposed technique over the conventional method of transmitting all the sensor data. As an example, state-of-the-art AFEs for ECG sensor typically consume 1 – 8× power of ADC [9], [11], [23], [63], and a

TABLE IV
COMPARISON WITH AI HARDWARE ACCELERATOR

	[13] VLSI'18	[14] JSSC'20	[15] ISSCC'19	[16] JSSC'18	[17] JSSC'20	[18] ISSCC'19	[41] TCASI'19	[19] ISSCC'21	[20] ISSCC'22	[42] DAC'22	This work
Computation type	SRAM					Analog					
	10T1C	12T	8T	6T	8T1C	ReRAM	ELM	3T DRAM	TD+SC	ROM	RC+ANN
Process	65nm	65nm	55nm	65nm	65nm	55nm	65nm	65nm	28nm	28nm	65nm
Weight precision	1	1	2	8	1	1	1	4	5	8	1(RC) 4-4-6(ANN) ¹
Input precision	1	1	1	8	1	3	8	4	5	8	12(RC) 10-8-8(ANN) ¹
Efficiency TOPS/W	658 ²	403 ²	18.4 ²	6.25 ²	671.5 ²	53.2 ²	2.9 ²	32.5 ²	13.3 ²	5.7 ²	43.6 ³
Norm. eff. (TOPS/W) ⁴	658	403	36.8	400	671.5	159.5	23.2	520	332.7	368	528 ³
Bitcell area(F ²) ⁵	7438.2	1300.2	189.3	577.8	1198.2	66.9	—	5847.8	1658.2	255.1	4890.1 ⁶

¹precision for 2 hidden layers and output layer; ²one MAC is considered as 2 OPS (multiplication and addition) and does not include energy for data movement and output activations; ³excludes output layer of RC; nonlinearity, ADC and DAC of RC are considered as 1 operation each; ⁴normalized efficiency is given by efficiency (TOPS/W) × input precision × weight precision; ⁵F: minimum feature size; ⁶corresponds to unit cell in switched-capacitor MAC circuit in the ANN

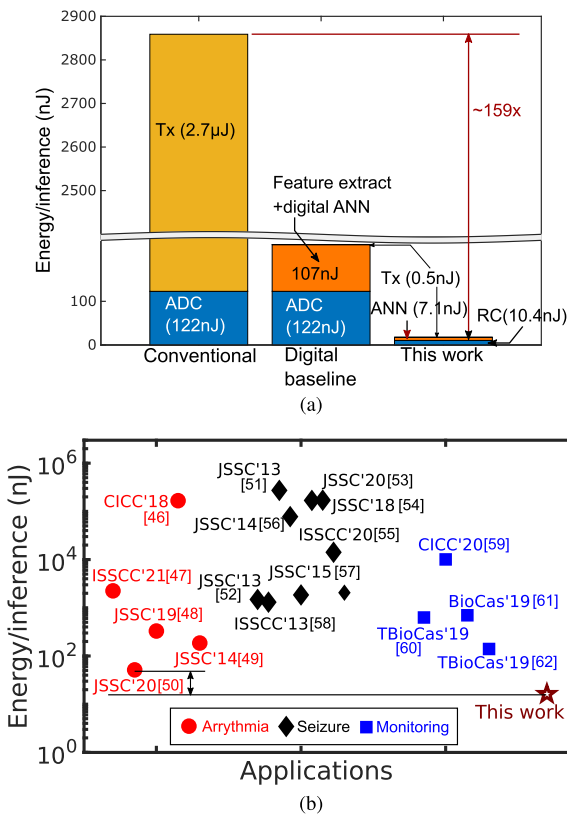


Fig. 12. Comparison with (a) conventional technique and baseline and (b) state-of-the-art AI ASICs for different bio-medical applications.

similar AFE in front of our RC+ANN will result in 3.8 – 21× reduction in energy compared to naive transmission. Hence, the in-sensor AI technique shifts the design burden for high energy efficiency from transmitter to the AFE. The AFE energy efficiency can be potentially improved through inverter based amplifier design and inverter stacking [64], [65] to reduce overall energy consumption of the sensor.

IV. CONCLUSION

This work has presented a fusion AI framework for sepsis prediction before onset, and in-sensor AI technique that can significantly reduce sensor energy by transmitting prediction scores instead of ECG samples. The energy consumption of the proposed in-sensor AI circuits can be reduced further by using dynamic amplifiers. Energy efficiency of the proposed AI circuits is expected to improve with technology scaling since the analog components do not need high linearity or gain.

ACKNOWLEDGMENT

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] J. Charpentier et al., "Brain natriuretic peptide: A marker of myocardial dysfunction and prognosis during severe sepsis," *Crit. Care Med.*, vol. 32, no. 3, pp. 660–665, 2004.
- [2] O. Court, A. Kumar, J. E. Parrillo, and A. Kumar, "Clinical review: Myocardial depression in sepsis and septic shock," *Crit. Care*, vol. 6, no. 6, pp. 1–9, 2002.
- [3] A. Rudiger and M. Singer, "Mechanisms of sepsis-induced cardiac dysfunction," *Crit. Care Med.*, vol. 35, no. 6, pp. 1599–1608, 2007.
- [4] M. M. Rich, M. L. McGarvey, J. W. Teener, and L. H. Frame, "ECG changes during septic shock," *Cardiol.*, vol. 97, no. 4, pp. 187–196, 2002.
- [5] J.-m. Kwon et al., "Deep learning model for screening sepsis using electrocardiography," *Scand. J. Trauma, Resuscitation Emerg. Med.*, vol. 29, no. 1, pp. 1–12, 2021.
- [6] A. M. R. Dixon, E. G. Allstot, D. Gangopadhyay, and D. J. Allstot, "Compressed sensing system considerations for ECG and EMG wireless biosensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 2, pp. 156–166, Apr. 2012.
- [7] C. J. Deepu, C.-H. Heng, and Y. Lian, "A hybrid data compression scheme for power reduction in wireless sensors for IoT," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 2, pp. 245–254, Apr. 2017.

- [8] W. Guo, Y. Kim, A. H. Tewfik, and N. Sun, "A fully passive compressive sensing SAR ADC for low-power wireless sensors," *IEEE J. Solid-State Circuits*, vol. 52, no. 8, pp. 2154–2167, Aug. 2017.
- [9] R. F. Yazicioglu, S. Kim, T. Torfs, H. Kim, and C. Van Hoof, "A 30 μ W analog signal processor ASIC for portable biopotential signal monitoring," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 209–223, Jan. 2011.
- [10] Y. Li, A. L. Mansano, Y. Yuan, D. Zhao, and W. A. Serdijn, "An ECG recording front-end with continuous-time level-crossing sampling," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 5, pp. 626–635, Oct. 2014.
- [11] S. Mondal, C.-L. Hsu, R. Jafari, and D. A. Hall, "A dynamically reconfigurable ECG analog front-end with a $2.5\times$ data-dependent power reduction," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 5, pp. 1066–1078, Oct. 2021.
- [12] M. Trakimas and S. R. Sonkusale, "An adaptive resolution asynchronous ADC architecture for data compression in energy constrained sensing applications," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 58, no. 5, pp. 921–934, May 2011.
- [13] H. Valavi, P. J. Ramadge, E. Nestler, and N. Verma, "A mixed-signal binarized convolutional-neural-network accelerator integrating dense weight storage and multiplication for reduced data movement," in *Proc. IEEE Symp. Very Large Scale Integration Circuits*, 2018, pp. 141–142.
- [14] S. Yin, Z. Jiang, J.-S. Seo, and M. Seok, "XNOR-SRAM: In-memory computing SRAM macro for binary/ternary deep neural networks," *IEEE J. Solid-State Circuits*, vol. 55, no. 6, pp. 1733–1743, Jun. 2020.
- [15] X. Si et al., "24.5 a twin-8T SRAM computation-in-memory macro for multiple-bit CNN-based machine learning," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2019, pp. 396–398.
- [16] S. K. Gonugondla, M. Kang, and N. R. Shanbhag, "A variation-tolerant in-memory machine learning classifier via on-chip training," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3163–3173, Nov. 2018.
- [17] Z. Jiang, S. Yin, J.-S. Seo, and M. Seok, "C3SRAM: An in-memory-computing SRAM macro based on robust capacitive coupling computing mechanism," *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1888–1897, Jul. 2020.
- [18] C.-X. Xue et al., "24.1 A 1Mb multibit ReRAM computing-in-memory macro with 14.6 ns parallel MAC computing time for CNN based AI edge processors," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2019, pp. 388–390.
- [19] Z. Chen, X. Chen, and J. Gu, "A 65 nm 3 T dynamic analog RAM-Based computing-in-memory macro and CNN accelerator with retention enhancement, adaptive analog sparsity and 44TOPS/W system energy efficiency," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2021, vol. 64, pp. 240–242.
- [20] J.-O. Seo, M. Seok, and S. Cho, "ARCHON: A 332.7 TOPS/W 5b variation-tolerant analog CNN processor featuring analog neuronal computation unit and analog memory," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2022, vol. 65, pp. 258–260.
- [21] S. Sadasivuni, M. Saha, N. Bhatia, I. Banerjee, and A. Sanyal, "Fusion of fully integrated analog machine learning classifier with electronic medical records for real-time prediction of sepsis onset," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022.
- [22] S. Mondal and D. A. Hall, "A 67- μ W ultra-low power PVT-Robust medradio transmitter," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, 2020, pp. 327–330.
- [23] L. Yan et al., "24.4 a 680nA fully integrated implantable ECG-acquisition IC with analog feature extraction," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2014, pp. 418–419.
- [24] F. C. Bauer, D. R. Muir, and G. Indiveri, "Real-time ultra-low power ECG anomaly detection using an event-driven neuromorphic processor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1575–1582, Dec. 2019.
- [25] K. Bai and Y. Yi, "DFR: An energy-efficient analog delay feedback reservoir computing system for brain-inspired computing," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 14, no. 4, pp. 1–22, 2018.
- [26] Y. Chen, E. Yao, and A. Basu, "A 128-channel extreme learning machine-based neural decoder for brain machine interfaces," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 3, pp. 679–692, Jun. 2016.
- [27] S. T. Chandrasekaran, S. P. Bhanushali, I. Banerjee, and A. Sanyal, "A bio-inspired reservoir-computer for real-time stress detection from ECG signal," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 290–293, 2020.
- [28] A. Rodan and P. Tino, "Minimum complexity echo state network," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 131–144, Jan. 2011.
- [29] S. T. Chandrasekaran, S. P. Bhanushali, I. Banerjee, and A. Sanyal, "Toward real-time, at-home patient health monitoring using reservoir computing CMOS IC," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 11, no. 4, pp. 829–839, Dec. 2021.
- [30] S. T. Chandrasekaran, A. Jayaraj, V. E. G. Karnam, I. Banerjee, and A. Sanyal, "Fully integrated analog machine learning classifier using custom activation function for low resolution image classification," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 68, no. 3, pp. 1023–1033, Dec. 2021.
- [31] A. Jalilifard, V. F. Caridá, A. F. Mansano, R. S. Cristo, and F. P. C. da Fonseca, "Semantic sensitive TF-IDF to determine word relevance in documents," in *Proc. Adv. Comput. Netw. Commun.*, 2021, pp. 327–337.
- [32] M. Nakhshabi, A. Toffy, P. Achuth, L. Palanichamy, and C. Vikas, "Early prediction of sepsis: Using state-of-the-art machine learning techniques on vital sign inputs," in *Proc. Comput. Cardiol.*, 2019, pp. 1–4.
- [33] A. D. Bedoya et al., "Machine learning for early detection of sepsis: An internal and temporal validation study," *JAMIA Open*, vol. 3, no. 2, pp. 252–260, 2020.
- [34] K. H. Goh et al., "Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare," *Nature Commun.*, vol. 12, no. 1, pp. 1–10, 2021.
- [35] C. Lin et al., "Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM," in *Proc. IEEE Int. Conf. Healthcare Inform.*, 2018, pp. 219–228.
- [36] R. Liu et al., "Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.
- [37] M. Medina and P. Sala, "On the early detection of sepsis in MIMIC-III," in *Proc. IEEE 9th Int. Conf. Healthcare Inform.*, 2021, pp. 171–180.
- [38] M. Rosnati and V. Fortuin, "MGP-AttTCN: An interpretable machine learning model for the prediction of sepsis," *Plos one*, vol. 16, no. 5, 2021, Art. no. e0251248.
- [39] S. Nemati et al., "An interpretable machine learning model for accurate prediction of sepsis in the ICU," *Crit. Care Med.*, vol. 46, no. 4, pp. 547–553, 2018.
- [40] D. A. Kaji et al., "An attention based deep learning model of clinical events in the intensive care unit," *PLoS one*, vol. 14, no. 2, Art. no. e0211057.
- [41] Y. Chen, Z. Wang, A. Patil, and A. Basu, "A 2.86-TOPS/W current mirror cross-bar-based machine-learning and physical unclonable function engine for internet-of-things applications," *IEEE Trans. Circuits Syst. I*, vol. 66, no. 6, pp. 2240–2252, Jun. 2019.
- [42] Y. Chen et al., "YOLoC: Deploy large-scale neural network by ROM-based computing-in-memory using residual branch on a chip," in *Proc. 59th IEEE/ACM Des. Automat. Conf.*, 2022, pp. 1093–1098.
- [43] Q. Dong et al., "15.3 a 351TOPS/W and 372.4 GOPS compute-in-memory SRAM macro in 7 nm FinFET CMOS for machine-learning applications," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2020, pp. 242–244.
- [44] P. P. Mercier et al., "A sub-nW 2.4 GHz transmitter for low data-rate sensing applications," *IEEE J. Solid-State Circuits*, vol. 49, no. 7, pp. 1463–1474, Jul. 2014.
- [45] B. Murmann, "ADC performance survey 1997–2021." [Online]. Available: <https://github.com/bmurmann/ADC-survey>
- [46] A. O. Akmandor, H. Yin, and N. K. Jha, "Simultaneously ensuring smartness, security, and energy efficiency in Internet-of-Things sensors," in *Proc. IEEE Custom Integr. Circuits Conf.*, 2018, pp. 1–8.
- [47] J. Liu et al., "A reconfigurable biomedical AI processor with adaptive learning for versatile intelligent health monitoring," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2021, vol. 64, pp. 62–64.
- [48] S. Yin et al., "A 1.06- μ W smart ECG processor in 65-nm CMOS for real-time biometric authentication and personal cardiac monitoring," *IEEE J. Solid-State Circuits*, vol. 54, no. 8, pp. 2316–2326, Aug. 2019.
- [49] S.-Y. Hsu, Y. Ho, P.-Y. Chang, C. Su, and C.-Y. Lee, "A 48.6-to-105.2 μ W machine learning assisted cardiac sensor SoC for mobile healthcare applications," *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 801–811, Apr. 2014.
- [50] K.-C. Chen, C.-Y. Chou, and A.-Y. Wu, "A tri-mode compressed analytics engine for low-power AF detection with on-demand EKG reconstruction," *IEEE J. Solid-State Circuits*, vol. 56, no. 5, pp. 1608–1617, May 2021.
- [51] 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 J. Solid-State Circuits*, vol. 48, no. 7, pp. 1625–1637, Jul. 2013.
- [52] J. Yoo, L. Yan, D. El-Damak, M. A. B. Altaf, A. H. Shueb, and A. P. Chandrakasan, "An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 214–228, Jan. 2013.
- [53] S.-A. Huang, K.-C. Chang, H.-H. Liou, and C.-H. Yang, "A 1.9-mW SVM processor with on-chip active learning for epileptic seizure control," *IEEE J. Solid-State Circuits*, vol. 55, no. 2, pp. 452–464, Feb. 2020.

- [54] H. Jia and N. Verma, "Exploiting approximate feature extraction via genetic programming for hardware acceleration in a heterogeneous microprocessor," *IEEE J. Solid-State Circuits*, vol. 53, no. 4, pp. 1016–1027, Apr. 2018.
- [55] Y. Wang, Q. Sun, H. Luo, X. Chen, X. Wang, and H. Zhang, "A closed-loop neuromodulation chipset with 2-level classification achieving 1.5Vpp CM interference tolerance, 35 dB stimulation artifact rejection in 0.5 ms and 97.8% sensitivity seizure detection," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2020, pp. 406–408.
- [56] W.-M. Chen et al., "A fully integrated 8-channel closed-loop neural-prosthetic CMOS SoC for real-time epileptic seizure control," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 232–247, Jan. 2014.
- [57] M. A. Bin Altaf, C. Zhang, and J. Yoo, "A 16-channel patient-specific seizure onset and termination detection SoC with impedance-adaptive transcranial electrical stimulator," *IEEE J. Solid-State Circuits*, vol. 50, no. 11, pp. 2728–2740, Nov. 2015.
- [58] M. A. B. Altaf, J. Tillak, Y. Kifle, and J. Yoo, "A 1.83 μ J/classification non-linear support-vector-machine-based patient-specific seizure classification SoC," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2013, pp. 100–101.
- [59] X. Tang and W. Tang, "A 151nW second-order ternary delta modulator for ECG slope variation measurement with baseline wandering resilience," in *Proc. IEEE Custom Integr. Circuits Conf.*, 2020, pp. 1–4.
- [60] A. Amirshahi and M. Hashemi, "ECG classification algorithm based on STDP and R-STDP neural networks for real-time monitoring on ultra low-power personal wearable devices," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1483–1493, Dec. 2019.
- [61] M. A. Sohail, Z. Taufique, S. M. Abubakar, W. Saadeh, and M. A. Bin Altaf, "An ECG processor for the detection of eight cardiac arrhythmias with minimum false alarms," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2019, pp. 1–4.
- [62] A. Dabbaghian, T. Yousefi, S. Z. Fatmi, P. Shafia, and H. Kassiri, "A 9.2-g fully-flexible wireless ambulatory EEG monitoring and diagnostics headband with analog motion artifact detection and compensation," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1141–1151, Dec. 2019.
- [63] P. Harpe, H. Gao, R. van Dommele, E. Cantatore, and A. van Roermund, "21.2 a 3nW signal-acquisition IC integrating an amplifier with 2.1 NEF and a 1.5 fJ/conv-step ADC," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, 2015, pp. 1–3.
- [64] L. Shen, N. Lu, and N. Sun, "A 1-V 0.25- μ W inverter stacking amplifier with 1.07 noise efficiency factor," *IEEE J. Solid-State Circuits*, vol. 53, no. 3, pp. 896–905, Mar. 2018.
- [65] Y.-P. Chen, D. Blaauw, and D. Sylvester, "A 266nW multi-chopper amplifier with 1.38 noise efficiency factor for neural signal recording," in *Proc. IEEE Symp. VLSI Circuits Dig. Tech. Papers*, 2014, pp. 1–2.



Sudarsan Sadasivuni (Member, IEEE) received the graduation degree from the University of Houston, Houston, TX, USA, in 2017. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering, University at Buffalo, Buffalo, NY, USA. His research interests include design of mixed signal machine learning circuits, data analysis, and developing deep learning network models for medical applications.



classification, computer-aided diagnosis, natural language processing, and cloud computing.



Sumukh Prashant Bhanushali received the B.Tech. degree from the K. J. Somaiya College of Engineering, Mumbai, India, and the M.S. degree from the University at Buffalo, Buffalo, NY, USA. He is currently working toward the Ph.D. degree electrical engineering with Arizona State University, Tempe, AZ, USA. He was an intern with Atom CPU Core Group, Intel, Austin, TX, USA, and with Alphacore, Tempe. His current research interests include design and validation of analog/mixed-signal circuits for machine learning, data converters, and image sensors.



Imon Banerjee received the M.Tech. degree from the National Institute of Technology, Durgapur, India, in 2011, and the Ph.D. degree from The University of Genova, Genova, Italy, in 2016. She did her post-doctoral training with Stanford University, Stanford, CA, USA. She was an Assistant Professor with the Department of Biomedical Informatics and Department of Radiology, Emory University, Atlanta, GA, USA, and an Assistant Professor with the Department of Biomedical Engineering, Emory University. She is currently the Lead AI Scientist with Mayo Clinic, Phoenix, AZ, USA, and an Associate Professor with the School of Computing and Augmented Intelligence, Arizona State University, AZ, USA. Her research interests include application of machine learning for biomedical data mining and predictive modeling. Dr. Banerjee was the recipient of 2012 Marie Curie fellowship in European FP7 Marie Curie Initial Training Networks.



Arindam Sanyal (Member, IEEE) received the B.E. degree from Jadavpur University, Kolkata, India, in 2007, and the M.Tech. degree from the Indian Institute of Technology, Kharagpur, India, in 2009, and the Ph.D. degree from The University of Texas at Austin, Austin, TX, USA, in 2016. He was an Assistant Professor with the Department of Electrical Engineering, University at Buffalo, Buffalo, NY, USA. He is currently an Assistant Professor with the School of Electrical, Computer and Energy Engineering, Arizona State University, AZ, USA. His research interests include analog/mixed signal design, bio-medical sensor design, analog security, and neuromorphic computing. Dr. Sanyal is an Associate Editor for *IET Electronics Letters*. He is a Member of Analog Signal Processing Technical Committee (ASP-TC) and VLSI Systems and Applications Technical Committee (VSA-TC) within the IEEE Circuits and Systems Society. He was the recipient of 2020 NSF CISE Research Initiation Initiative (CRII) Award, Intel/Texas Instruments/Catalyst Foundation CICC Student Scholarship Award, 2014 and Mamraj Agarwal Award in 2001.