DNN-Assisted Sensor for Energy-Efficient ECG Monitoring

Tao-Yi Lee, Marco Levorato, Nikil Dutt
Department of Computer Science
University of California, Irvine
Irvine, USA
{taoyil, levorato, dutt}@uci.edu

***Abstract*—The quasi-periodic nature of electrocardiogram (ECG) signals enables the use of compression techniques to minimize communications and reduce energy intake for diagnostic and preventive health monitoring. However, compression often degrades signal quality and may impair analysis by means of machine learning algorithms for the detection of anomalies. In this paper, we present an approach to pre-select relevant portions of the ECG signal at the sensor to reduce network load while satisfying a predefined diagnostic sensitivity requirement. We deploy a Deep Neural Network (DNN) to filter-out the signal's normal rhythms and reduce the amount of data stored or transmitted for further processing. Our extensive experiments covering a wide range of DNN hyper-parameters illustrate the trade-off between diagnostic sensitivity, channel usage, energy consumption and computational complexity.**

***Index Terms*—electrocardiogram, signal compression, convolutional neural networks**

I. Introduction

After more than a century since it was first invented, the electrocardiogram (ECG) has become a commodity cardiovascular disease (CVD) diagnostic tool. Nowadays, the acquisition of ECG signals can be performed by wearable devices [1]. The availability of ECG signals enables a broad spectrum of applications, ranging from early detection of premonitory signs of myocardia ischmia and arrhythmia to long term monitoring for the identification of rare but life-threatening rhythms. Recent contributions by Hannun et. al. [2] have demonstrated cadiologist-level detection and classification of arrhythmia by means of specialized convolutional neural networks (CNN).

One existing clinical device, the Holter monitor, allows continuous acquisition of ECG signals for extended time periods. However, due to the shear amount of data, expert interpretation of Holter readings is often performed offline, sometimes severely delayed. Furthermore, some features in ECG recording which suggest heart malfunction only appear infrequently. For instance, premature ventricular contractions (PVCs) is an elusive rhythm which might lead to the diagnosis of heart failure (HF) and cardiomyopathy. According to [3], 24-hour or 48-hour Holter monitoring is usually prescribed to identify a key determinant of PVC – ventricular premature beats (VPB) – when a patient complains of dizziness or palpitations after there were no specific findings in routine 12-lead ECG. Events of VPB might occur as infrequently as once a week or even a couple of times a month [4]. Naively transmitting and recording complete ECG signals seems to be wasteful in dealing with such rare rhythms.

Therefore, we assume there exists a way to efficiently collect and deliver ECG signals to the final analysis unit without employing a high-capacity communication channel. In other words, current wireless transceivers have potential to support more sensors without any increase of the energy budget.

ECGs are amenable to compression techniques due to their quasi-periodicity, but compression must be applied carefully to avoid any distortion of features instrumental to classification. In fact, while Deep Neural Networks (DNN) appear to be an ideal tool to analyze ECG signals at the clinical level, this family of tools requires a high precision representation of the full morphology of the signal. Several works have explored compression techniques to mitigate this bandwidth requirement. The most relevant to the systems investigated herein is *compressive sensing* (CS) [5–7]. CS exploits the repetitive nature of ECG and employs l_1-minimization to approximate raw signals using a sparse reconstruction matrix with minimum distortion. To obtain sparse representation of ECG, wavelet transform[8] and dictionary learning[9] are commonly adopted. Other approaches to obtain a high-fidelity signal recovery, such as auto-encoders [10], are often too complex to be executed in-sensor.

The main limitation of traditional compression approaches is that they aim to generate a high-quality reconstruction of the entire signal. This necessarily results in a large amount of storage and/or channel capacity needed to represent normal rhythms, that may be useless for the final diagnosis objective, which only requires anomalous rhythms. Here, we take the different approach of "pre-selecting" interesting and potentially anomalous portions of the ECG signals, by completely removing uninformative normal heart cycles from the signal representation. To this aim, we train a deep neural network (DNN) that is executed at the sensor and which provides as output the probability that a rhythm is anomalous. A threshold strategy is then used to determine whether or not the rhythm

This work was partially supported by the US National Science Foundation (NSF) WiFiUS grant CNS-1702950. Data used for this research was provided by the Telemetric and Holter ECG Warehouse of the University of Rochester (THEW), NY. The authors would also like to thank Su-Han Chang, M.D. for her input on the clinical aspects of electrocardiogram.

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

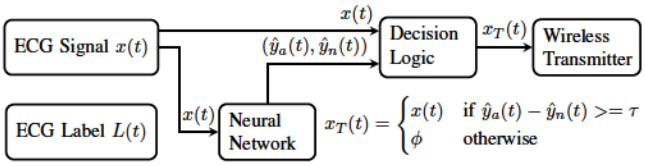


Fig. 1. Schematics of the proposed system.

is transmitted and stored. Intuitively, the execution of complex DNNs, such as those providing fine-grain classification of ECG cycles, are likely infeasible on wearable sensors. Thus, herein we provide a thorough evaluation of a wide-range of DNN parameters to explore the tradeoff between complexity/energy consumption, channel usage and diagnosis accuracy.

The rest of the paper is organized as follows. Section II begins by formulating a quantitative framework that evaluates how well the system performs in terms of mistransmission rate and normalized energy cost. In Section III, we present a detailed description of the proposed system and its critical functional blocks. In Section V, we present experimental results to numerically validate the proposed methods using a parameterized CNN energy estimation model. Section VI concludes the paper.

II. Problem Formulation

We consider a scenario where a sensor acquires an ECG signal and wirelessly transmits it to a server. Additionally, we assume that *only* the abnormal rhythms need to be transmitted for further interpretation by either an expert or a machine learning model. Intuitively, transporting the whole signal to a server does not require any computation at the sensor, but imposes a high traffic load to the wireless channel, whose capacity is often limited. Compression can reduce traffic. However, compression in narrow sense assumes lossless recovery of all the rhythms whose completeness may not be necessary in the application. Arguably, adopting *lossy* compression can aggressively reduce channel usage while taking a toll on the overall quality of recovered signal and the final classification accuracy on the server end.

The technique we propose whose schematic is represented in Fig. 1 takes the different approach of pre-selecting at the sensor abnormal rhythms. A neural network is followed by a decision logic whose objective is to discard all rhythms deemed normal. Clearly, the sensor cannot afford to execute the full-fledged classifier, whose use would result in no accuracy loss and minimal channel usage. We, then, explore the tradeoff between the computational complexity pushed to the sensor, and the performance of the system. We explicitly consider four metrics: (a) overall accuracy of the classification process, (b) bandwidth utilization, (c) pre-selection accuracy, and (d) energy usage.

Formal definitions of the signals and performance metrics mentioned earlier are provided below. We consider signal-label pairs consisting of a digitized ECG signal and sample-wise anomaly labels annotated by ECG specialists:

$$x(t) = [x_1 \ , \ x_2 \ , \ x_3 \ , \ x_4 \ , \ x_5 \ , \ldots, \ x_n \ ] \quad (1)$$
$$L(t) = [(1,0), \ (0,1), \ (1,0), \ (0,1), \ (0,1), \ldots, \ (1,0)] \quad (2)$$

where brackets denote a time series of discrete samples. Labels in the following discussion are expected to take a 2-tuple format where abnormality and normality are indicated by the first and the second element, respectively. For abnormal samples, say x_1 in Eq. 3, the label is $(1,0)$. On the other hand, $(0,1)$ are tagged on normal samples such as x_3. Note that not all samples are annotated. Therefore, some samples may have $(0,0)$ meaning they are neither tagged positive nor negative. Acquisition is followed by decision logic, denoted by $\mathcal{D}\{\cdot\}$, which selects positive samples from $x(t)$ and output transmitting time series $x_T(t)$. For example, if samples x_2, x_4 and x_5 are predicted to be associated with normal rhythms, $\mathcal{D}\{\cdot\}$ replaces them with a special symbol ϕ to command wireless transmitter *not* to transmit those samples. The following equation shows the corresponding output of decision logic if the realization of $x(t)$ in Equation (1) is the input:

$$x_T(t) = \mathcal{D}\{x(t)\} = [x_1, \ \phi, \ x_3, \ \phi, \ \phi, \ldots, \ x_n]. \quad (3)$$

Intuitively, the bandwidth utilization b_{tx} of a ECG frame is defined as ratio of valid samples between raw signal $x_T(t)$ and transmitted signal $x(t)$

$$b_{tx} = \frac{|\{x_i : x_i \in x_T(t) \wedge x_i \neq \phi\}|}{|\{x_i : x_i \in x(t)\}|}. \quad (4)$$

To evaluate the loss of sensitivity caused by dropped positive samples, we use the ground truth $L(t)$ to define the mistranmission rate ϵ_{mt} as the ratio of positive samples between $x_T(t)$ and in $x(t)$, that is,

$$\epsilon_{mt} = \frac{|\{x_i : x_i \in x_T(t) \wedge L(t_i) = (1,0)\}|}{|\{x_i : x_i \in x(t) \wedge L(t_i) = (1,0)\}|}. \quad (5)$$

A. Energy Consumption

To justify the new system architecture, we raise the following question: *Will the energy savings in RF transmission be greater than the cost of running an extra CNN?*

To address this question, we propose in this section an analytical model for energy consumption that takes into consideration both RF transmission and CNN inference[11].

TABLE I
Normalized energy costs for critical operations in unified energy model (assume a and b are 8-bit numbers)

Operation	Symbol	Norm. Energy	Energy
DRAM R/W	E_d	4000 ×	320 pJ
Global Buffer R/W	E_g	15 ×	1.2 pJ
Inter-PE Comm.	E_p	5 ×	400 fJ
Register File R/W	E_r	2.5 ×	200 fJ
$a + b$	E_a	10^{-2} ×	0.8 fJ
$a \times b$	E_m	1 ×	80 fJ
RF Transmission	E_{tx}	10^6 ×	80 μJ

If we consider the RF transmitter energy model described in [12], transmitter energy consumption E_{tx} can be broken down into two components: transmitted energy E_t, and dissipated energy E_h, *i.e.* waste heat, as in

$$E_{tx} = E_t + E_h = (P_t + P_h)T_{tx} = P_{tx}T_{tx} \tag{6}$$

where T_{tx} is the accumulated time for active signaling, *i.e.* $x_T(t) \neq \phi$. Since we are targeting a body area network (BAN) application, we set the transmitter-receiver distance to $d = 10m$. For simplicity, we select 8-DPSK, which transports 3 bits per symbol, as the modulation scheme. Transmission energy cost is approximately 10 μJ/bit, if we assume the same set of system design parameters as in [12].

In order to estimate the energy consumption per inference, we used an energy model analogous to proposed by Chen et. al. in [13]. The energy cost associated with each operation are listed in Table I where we assume 8-bit fixed-point numbers in each operation. Then, the average energy $\mathcal{E}$ for the proposed system to process an ECG sample can, then, be computed as

$$\begin{aligned}\mathcal{E} \approx l_f\big(& (\gamma_{wd}^{-1}\vec{1}_l \cdot \vec{n}_c + \gamma_{dd}^{-1})E_d \\ & + (\gamma_{wg}^{-1}\vec{1}_l \cdot \vec{n}_c + \gamma_{dg}^{-1})E_g \\ & + (\gamma_{wp}^{-1}\vec{1}_l \cdot \vec{n}_c + \gamma_{dp}^{-1})E_p \\ & + (\gamma_{wr}^{-1}\vec{1}_l \cdot \vec{n}_c + \gamma_{dr}^{-1})E_r \\ & + \vec{1}_l \cdot \vec{n}_c(E_a + E_m)\big) + b_{tx}E_{tx} = E_{comp} + b_{tx}E_{tx}\end{aligned} \tag{7}$$

where $\cdot$ denotes vector inner product, l_f is the length of filter; $\vec{1}_l$ denotes an all one vector of length l; $\vec{n}_c$ is the number of channels in the CNN. Data and weight reuses are modeled by γ factors: the first character in subscript denotes data (d) and weight (w), while the second subscript denotes a particular memory hierarchy. Specifically γ_{dd}, γ_{dg}, γ_{dp}, and γ_{dr} corresponds to data reuse at DRAM (d), global buffer (g), array (p) and register files (r), respectively. Note Eq. 7 represents the energy cost to generate a single output.

The relative reduction in energy over a naive system is

$$\Delta\mathcal{E} = \frac{\mathcal{E}_{naive} - \mathcal{E}}{\mathcal{E}_{naive}} = \frac{E_{tx} - \mathcal{E}}{E_{tx}}, \tag{8}$$

a metric which we will use in Section V to assess the savings in energy consumption through our proposed approach.

III. System Architecture

In the following, we describe the components of the system in detail.

A. Neural Network

Intuitively, a neural network as complex as that used at the remote device to perform final classification would enable a "perfect" selection of anomalous signal sections. In fact, as the two classifiers have the same output, only positive cycles would be selected for transmission, thus minimizing channel usage without any accuracy degradation. Clearly, such a complex neural network cannot be executed within wearable sensors. Therefore we explore a range of parameters corresponding to different levels of complexity – and thus energy consumption and selection accuracy – for DNNs deployed on wearable sensors.

We trained a total of 26 variants of customized sequence to sequence CNN (S2SCNN) models to translate single lead ECG signals into sample-wise annotation of arrhythmia. Fig. 2 shows a representative version, which has $800,010$ parameters, composed of 4 convolution (CONV) modules. The output of the CNN consists of two probability estimates $\hat{y}_a$, and $\hat{y}_n$, indicating probabilities of abnormality and normality, respectively. Each CONV module consists of a 64-channel 1D convolution filter which has a filter length of 64, a batch normalization layer, rectified linear (ReLU) activation layer, and a dropout layer set to a 25% dropout rate. Since the training data only has labels marked on QRS complexes, at the output of each network, a softmax layer is employed to enable generation of the most decisive predictions by the CNN.

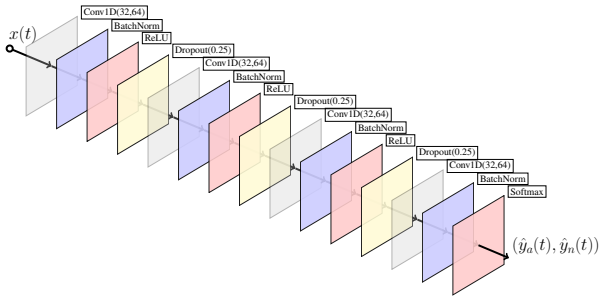


Fig. 2. Two representative architectures of CNN employed in this work. Layer caption "Conv1D(α, β)" denotes a 1D convolution layer which has α filter kernels (channels) and β filter length; "Dropout(γ)" denotes dropout[14] layer with rate of γ.

All CNNs are implemented in Keras with Tensorflow backend. Training is performed de novo with random initialization of weights. We used the Adam optimizer with gradient clipping enabled and a mini batch size of 128. The initial learning rate is set to 1×10^{-3} and reduced by tenfold when validation loss levels off. Convergence can be reached within 20 training epochs in half an hour on the nVidia 2080Ti GPU.

B. Decision Logic

Signal compression is realized by the decision logic, which takes the raw signal $x(t)$ and the outputs of the neural network to determine whether sections of the signal are transmitted or not. The logic will selectively transmit samples if the output of the neural network exceeds threshold τ. Transfer function of the decision logic can be formulated by

$$x_T(t) = \begin{cases} x(t) & \text{if } \hat{y}_a(t) - \hat{y}_n(t) >= \tau \\ \phi & \text{otherwise.} \end{cases} \tag{9}$$

Threshold τ is a system design parameter which allows adjustment on sensitivity. Lower τ is more conservative in generating negative predictions. Also, reducing τ implies more data will be transmitted.

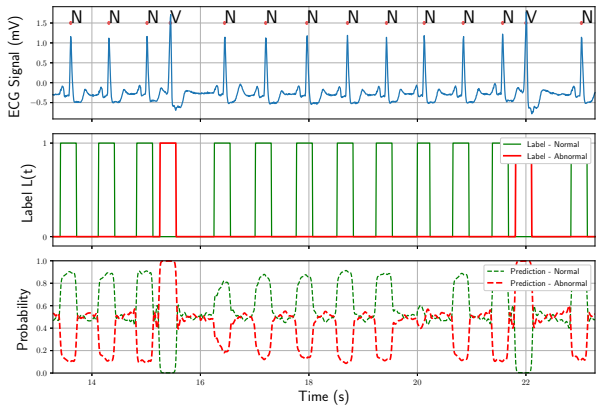


Fig. 3. Inference results on MITDB record number 105 via CNN which has 137,510 parameters. Blue curve is the Modified lead II (MLII) signal. Note that two VPC events, indicated by symbol "V", are observed in this interval. Middle panel shows the label which is derived from annotation files; Bottom panel shows prediction outputs for normal $\hat{y}_n(t)$ and abnormal heartbeats $\hat{y}_a(t)$

TABLE II
PERFORMANCE SUMMARY OF CNN VARIANTS. AUC: AREA UNDER ROC CURVE; TPR: TRUE POSTIVE RATE WHEN $\tau = 0$.

Parameters	#CONV	Channels $\vec{n}_c$	AUC	TPR
1460160	4	[106, 106, 106, 2]	0.862	0.683
1224441	4	[97, 97, 97, 2]	0.863	0.799
1032321	4	[89, 89, 89, 2]	0.868	0.729
856585	4	[81, 81, 81, 2]	0.867	0.782
800010	5	[64, 64, 64, 64, 2]	0.849	0.737
697233	4	[73, 73, 73, 2]	0.864	0.759
537546	4	[64, 64, 64, 2]	0.870	0.742
413010	4	[56, 56, 56, 2]	0.880	0.759
343495	4	[51, 51, 51, 2]	0.863	0.762
330360	4	[50, 50, 50, 2]	0.870	0.779
317481	4	[49, 49, 49, 2]	0.860	0.777
304858	4	[48, 48, 48, 2]	0.867	0.813
280380	4	[46, 46, 46, 2]	0.845	0.575
268525	4	[45, 45, 45, 2]	0.869	0.676
256926	4	[44, 44, 44, 2]	0.864	0.753
245583	4	[43, 43, 43, 2]	0.855	0.778
234496	4	[42, 42, 42, 2]	0.844	0.759
213090	4	[40, 40, 40, 2]	0.882	0.795
137706	4	[32, 32, 32, 2]	0.874	0.724
91920	4	[26, 26, 26, 2]	0.864	0.769
72010	3	[32, 32, 2]	0.865	0.836
60805	4	[21, 21, 21, 2]	0.863	0.731
48526	3	[26, 26, 2]	0.877	0.716
36090	4	[16, 16, 16, 2]	0.870	0.724
32476	3	[21, 21, 2]	0.842	0.819
19626	3	[16, 16, 2]	0.858	0.719

IV. EXPERIMENTAL SETUP

A. Dataset

To evaluate the proposed system, we aim to maximize energy efficiency in recording *arrhythmia* which is evident from uneven heartbeats in the ECG. We obtained multiple ECG datasets from PhysioNet and Telemetric and Holter ECG warehouse (THEW)[15]. Five specific datasets are combined to obtain a 205-record dataset for CNN training and system evaluation. Each of the records has associated labeling information which is compliant to ANSI/AAMI EC57 standard (AAMI). The dataset is then split into training, development, and test subsets. Each of them has 184, 10 and 11 records, respectively. Records from the same subject do not overlap or duplicate across different datasets. According to Luz et. al.[16], (modified) lead II (II/MLII) is selected to be the most informative feature for arrhythmia detection.

Abnormal rhythms are derived from disjunction of S, V, F, and Q categories; Normal rhythms are marked by label N. Indexes of both types of heartbeats are converted into two impulse trains $\delta_a(t)$ and $\delta_n(t)$, respectively. To account for average PR interval and QRS complex duration, the intervals advancing and lagging R-peak 0.2 seconds and 0.1 seconds are both assumed to share the same label assigned to corresponding R-peak.

TABLE III
RELATIVE ENERGY SAVINGS ON DIFFERENT RECORDS IF 1% ϵ_{mt} IS GUARANTEED

Record	$\Delta\mathcal{E}$ (%)	τ	b_{tx}(%)
105	33.91	-0.14	59.71
1121a	18.12	0.03	75.50
210	13.47	-0.44	80.15
114	8.48	-0.56	85.15
ECG-P18-02	-3.65	-0.72	97.28
107	-4.75	-0.14	98.37
124	-5.65	-0.81	99.28
Mean	8.56	-0.40	85.06
Max	33.91	-0.03	99.28
Min	-5.65	-0.81	59.71

B. Methodology to select system parameters

Three system parameters, namely 1) topology of CNN $\boldsymbol{p}$, 2) computation resource parameters $\boldsymbol{\gamma}$, and 3) communication system parameters $\boldsymbol{t}$, determine the overall system performance which is defined by application performance ϵ_{mt} and energy cost $\mathcal{E}$. Here we assume 2) and 3) are fixed and orthogonal to the CNN parameters. On dataset $\mathcal{D}$, the problem of finding the optimal system parameters can be formulated as

$$\hat{\boldsymbol{p}} = \underset{\boldsymbol{p}\in\mathcal{P}}{\text{argmin}}\ r_p[\text{CNN}_{\boldsymbol{p}}(\mathcal{D})]P_{tx,\boldsymbol{t}} + \sigma[\text{CNN}_{\boldsymbol{p}}(\mathcal{D})]E_{comp,\boldsymbol{\gamma}}$$
$$\text{subject to } \epsilon_{mt} < T, \tag{10}$$

where $r_p[\text{CNN}(\mathcal{D})]$ and $\sigma[\text{CNN}(\mathcal{D})]$ denote the rate of positive calls of CNN and the total amount of computations, respectively. We applied grid search on parameter space $\mathcal{P}$ to determine the best configuration of $\boldsymbol{p}$ which is derived numerically in the next section.

V. NUMERICAL RESULTS

A. Performance Summary

Table III summarizes the energy savings as computed in Eq. 8. Only the best performing CNN (indicated by a * in Fig. 4) is represented in the table while simulations are performed

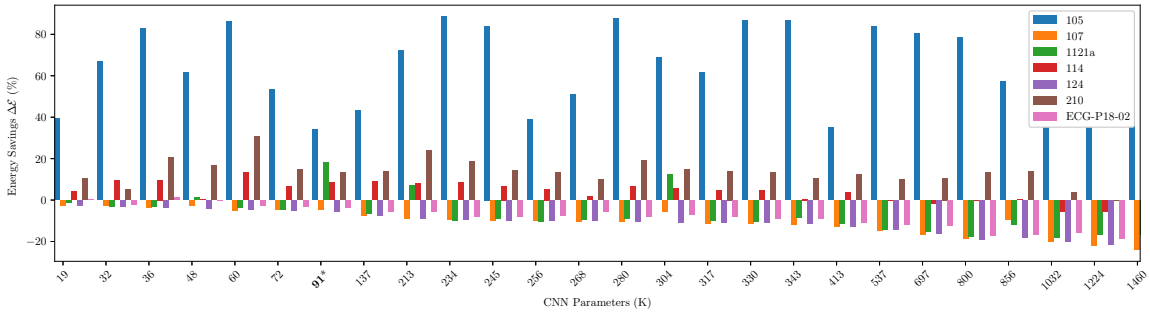
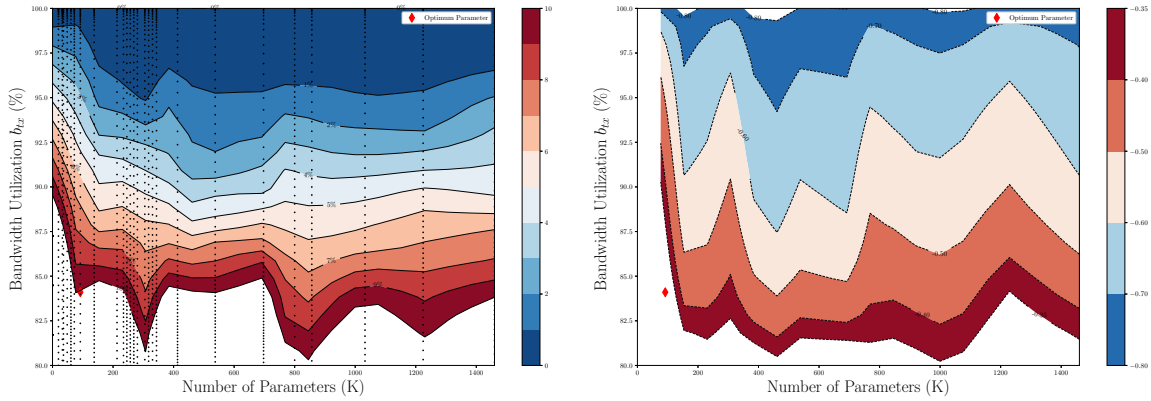


Fig. 4. Achievable energy saving among all CNN



(a) Contour plot of mistransmission rate (%)

(b) Contour plot of decision threshold τ

Fig. 5. Mistransmission rate among 6 test records. Real data points are indicated by black dots. Two subplots should be read relationally. Any single point on the left contour has a corresponding point on the right. Lower threshold value τ (color coded blue) corresponds to a more conservative DNN which leads to more data transmission.

based on all variants of CNNs listed in Table II. The threshold is set to allow at most 1% of mistransmission rate ϵ_{mt}.

Contour plots of the mistransmission rate are shown in Fig. 5a with corresponding threshold τ in Fig. 5b. The contours are triangularized on mesh grids of b_{tx} and number of parameters in the CNN. Note that setting the threshold to -1 results in $\epsilon = 0\%$ as all samples are transmitted. On the ordinate, where the number of parameters is zero, data points are computed based on "stochastic transmitter" assumption. The stochastic transmitter is defined to randomly drop samples for bandwidth saving. Asymptotic limit of the mistransmission rate when the number of parameters approaches zero, *i.e.* no CNN in system, is estimated under this assumption.

B. Optimal system parameters

As summarized in Fig. 4, the overall best performance can be achieved with the CNN which has 91,920 parameters. This result is consistent with the highlighted design point closest to the origin in Fig. 5a. The energy saving is a direct result of the CNN performance.

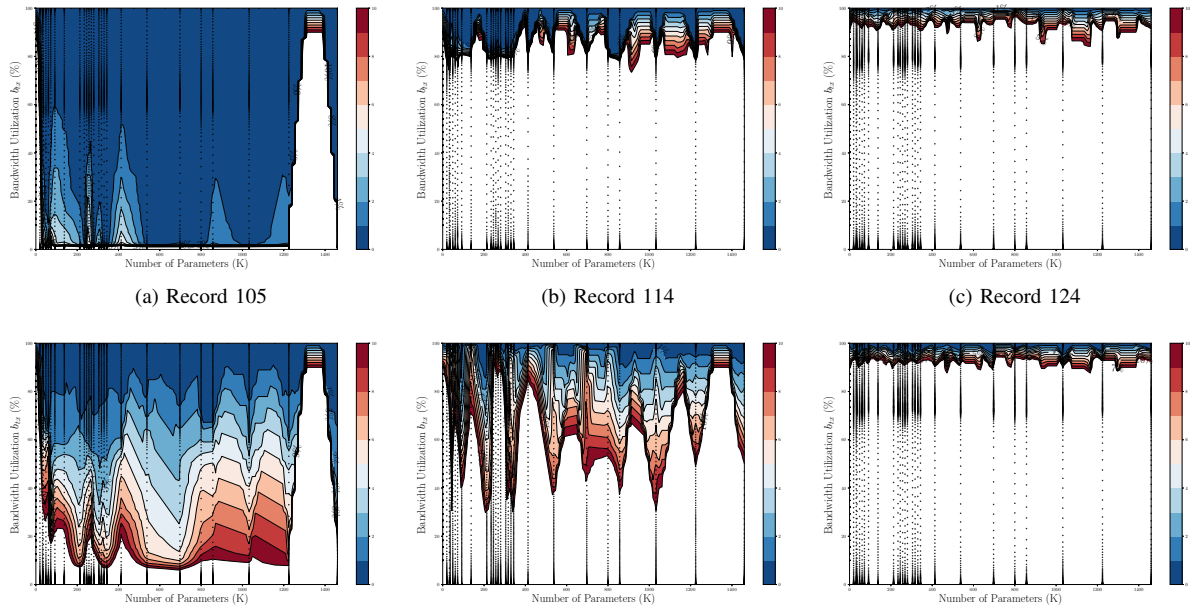
VI. Conclusions

We proposed a new technique that leverages a lightweight DNN to minimize energy consumption on wearable ECG sensors by selectively transmitting only the informative parts of an acquired ECG signal. We also formulated a unified energy model for DNN-assisted embedded systems. Our extensive experiments on real world datasets demonstrate that our proposed methodology locates the optimal CNN hyperparameters that minimizes the overall energy cost, while maintaining classification performance above a predefined threshold, making this a viable approach for deploying DNNs on resource constrained sensors.

References

[1] A. Bansal and R. Joshi, "Portable out-of-hospital electrocardiography: A review of current technologies," *Journal of Arrhythmia*, vol. 34, no. 2, pp. 129–138, Feb. 23, 2018.

[2] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, Jan. 2019.



(a) Record 105 (b) Record 114 (c) Record 124

(d) Record 210 (e) Record 1121a (f) Record ECG-P18-02

Fig. 6. Mistransmission per record.

[3] R. J. Simpson, W. E. Cascio, P. J. Schreiner, R. S. Crow, P. M. Rautaharju, and G. Heiss, "Prevalence of premature ventricular contractions in a population of African American and white men and women: The Atherosclerosis Risk in Communities (ARIC) study," *American Heart Journal*, vol. 143, no. 3, pp. 535–540, Mar. 2002.

[4] R. G. Hiss and L. E. Lamb, "Electrocardiographic Findings in 122,043 Individuals," *Circulation*, vol. 25, no. 6, pp. 947–961, Jun. 1962.

[5] L. F. Polania, R. E. Carrillo, M. Blanco-Velasco, and K. E. Barner, "Compressed sensing based method for ECG compression," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 761–764.

[6] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed Sensing for Real-Time Energy-Efficient ECG Compression on Wireless Body Sensor Nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2466, Sep. 2011.

[7] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, and E. Jones, "Energy-efficient Compressed Sensing for ambulatory ECG monitoring," *Computers in Biology and Medicine*, vol. 71, pp. 1–13, Apr. 2016.

[8] and and W. A. Pearlman, "Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 849–856, Jul. 2000.

[9] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary Learning Algorithms for Sparse Representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, Feb. 2003.

[10] D. Del Testa and M. Rossi, "Lightweight lossy compression of biometric patterns via denoising autoencoders," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2304–2308, 2015.

[11] B. Martinez, M. Monton, I. Vilajosana, and J. D. Prades, "The Power of Models: Modeling Power Consumption for IoT Devices," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5777–5789, Oct. 2015.

[12] F. Mahmood, E. Perrins, and L. Liu, "Modeling and Analysis of Energy Consumption for RF Transceivers in Wireless Cellular Systems," in *2015 IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, USA: IEEE, Dec. 2015, pp. 1–6.

[13] Y. Chen, J. Emer, and V. Sze, "Eyeriss: A Spatial Architecture for Energy-Efficient Dataflow for Convolutional Neural Networks," in *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Jun. 2016, pp. 367–379.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," p. 30,

[15] S. Kaab, "Sotalol testing unmasks altered repolarization in patients with suspected acquired long-qt-syndrome: A case-control pilot study using i.v. sotalol," *European Heart Journal*, vol. 24, no. 7, pp. 649–657, Apr. 2003.

[16] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144–164, Apr. 2016.