

BiometricNet: Deep Learning based Biometric Identification using Wrist-Worn PPG

Luke Everson^{1,2}, Dwaipayan Biswas¹, Madhuri Panwar³, Dimitrios Rodopoulos¹, Amit Acharyya³, Chris H. Kim²,
Chris Van Hoof¹, Mario Konijnenburg⁴, Nick Van Helleputte¹

¹IMEC, Heverlee, Belgium, {Dwaipayan.Biswas, Dimitrios.Rodopoulos, Chris.VanHoof, Nick.VanHelleputte}@imec.be

²University of Minnesota, Minneapolis, Minnesota, USA {evers193, chriskim}@umn.edu

³Indian Institute of Technology Hyderabad, Telangana, India {ee15resch01004, amit_acharyya}@iiith.ac.in

⁴Holst Centre, Eindhoven, Netherlands (Mario.Konijnenburg@imec-nl.nl)

Abstract— Rapid advances in semiconductor fabrication technology have enabled the proliferation of miniaturized body-worn sensors capable of long term pervasive biomedical signal monitoring. In this paper, we present a novel deep learning-based framework (BiometricNET) on biometric identification using data collected from wrist-worn Photoplethysmography (PPG) signals in ambulatory environments. We have formulated a completely personalized data-driven approach, using a four-layer deep neural network - employing two convolution neural network (CNN) layers in conjunction with two long short-term memory (LSTM) layers, followed by a dense output layer for modelling the temporal sequence inherent within the pulsatile signal representative of cardiac activity. The proposed network configuration was evaluated on the TROIKA dataset collected from 12 subjects involved in physical activity, achieved an average five-fold cross-validation accuracy of 96%.

Keywords— *biometric, PPG, deep learning, convolutional neural network, long short-term memory.*

I. INTRODUCTION

Biometrics as defined by the International Organization for Standardization (ISO) is the “automated recognition of individuals based on their behavioral and biological characteristics” where common characteristics include fingerprints, iris, DNA, voice and gait analysis [1]. Recent studies on biometrics have also focused on one-dimensional physiological signals, i.e. electrocardiograms (ECG) [2], electroencephalograms (EEG) [3], phonocardiogram (PCG) and lastly photoplethysmography (PPG) [4]. These physiological signals provide an insight into the clinical condition of subjects, hence paving the way for personalization and identification of individuals. Such signals also have the distinct advantage of enabling continuous authentication systems since they can be captured for long durations without manual intervention. While ECG is more established and robust for identification than PPG, it requires multiple electrodes placed on the chest making it inefficient in terms of wearability for daily life usage. PPG signals are obtained from pulse oximeters which emit light on the skin and measure the change of light intensity, which is either transmitted or reflected through the skin. The periodicity of the reflected/transmitted light corresponds to the cardiac rhythm, often used for heart rate estimation. This makes PPG sensors a popular choice for embedding into wearable devices (e.g. smart watches). PPG signals can be acquired from various positions such as earlobes, fingertips or wrist, with the latter considered

as a convenient position for unobtrusive daily use. However, the acquisition is vulnerable to motion artifacts (MA) in normal daily living conditions and correspondingly distorts the signal fidelity and inhibit identification robustness.

PPG-based biometric identification has been a well-researched topic, but the majority of them have used signals collected in clinical settings which are less prone to MA, thereby, making these models unsuitable for daily life usage, given the fact that motion artifacts are ever present during daily activity of the user [3]. Earlier efforts on PPG-based identification have focused on data collected from the finger and analyzed using frequency domain analysis (Fourier analysis), correlation, peak detection, feature extraction and classification, employing fuzzy-logic [3-4] and Linear Discriminant Analysis [5], yielding high accuracies, approximately between 90-95%. Furthermore, approaches based on predictive learning methods relied on hand crafted features with some research extracting up to 40 features used in conjunction with a k-nn classifier [6]. On this backdrop, a recent work has focused on a two-stage procedure involving clustering (using 11 hand-crafted features) and deep learning techniques (Restricted Boltzmann Machines and Deep Belief Networks) [7]. It is evaluated on PPG signals collected from the wrist using green light under ambulant conditions. Results have been promising which paves the way for future research involving no feature extraction and entirely relying on the neural network for subject identification. Furthermore, it would also incur a research challenge on the use-case of such a system involving deployment, model update based on the changing cardiac and physical conditions of the subjects and hardware security and privacy which are yet to be explored on a system level.

In this paper, our key contribution lies in the fact that we use a completely data-driven approach based on convolution neural network (CNN) in conjunction with long and short term memory (LSTM) for modelling the underlying temporal sequence in the biological data of each subject, which negates the requirements of data processing and heuristics involved in popularly used classification schemes. Our proposed framework, *BiometricNET*, has been motivated by the fundamentals of Deep Neural Network (DNN), since it obviates the need for feature engineering, besides being successful in a wide range of applications. The paper is further structured as follows. Section II describes the problem formulation using our learning-based approach. The proposed methodology highlighting the DNN fundamentals and the

This research was supported in part by NSF IGERT grant DGE-1069104.

developed network architecture, *BiometricNET* have been detailed in section III. The results have been presented in section IV and the conclusions have been drawn in section V.

II. PROBLEM FORMULATION

In this paper, we adopt a completely personalized approach using DNN for a robust biometric identification. For this exploration we evaluate our algorithm on the IEEE Signal Processing Cup database comprising of PPG signals from 12 healthy male subjects, age ranging from 18 to 35 [8]. The subjects walked or ran on a treadmill with the following speeds in order: 1–2 km/h for 0.5 min, 6–8 km/h for 1 min, 12–15 km/h for 1 min, 6–8 km/h for 1 min, 12–15 km/h for 1 min, and 1–2 km/h for 0.5 min. Each subject's data contains two channels of PPG, recorded from the wrist (dorsal) using a pulse oximeter with green LED (wavelength: 515 nm); tri-axial accelerometer signals also recorded from the wrist, and a channel of ECG recorded from the chest using wet ECG electrodes, all recorded simultaneously. The ECG signal is the source of ground-truth for heart rate extraction from corresponding PPG signals, but is not required or used in our exploration. All signals were sampled at 125 Hz and transmitted to a computer through Bluetooth.

We formulate the problem as a binary classification task (one vs all) wherein a given subject is identified against a group of subjects based on the individual's PPG. For training the network, we have a significantly unbalanced class distributions, given that we have 12 subjects, the imbalance is 11:1. During training, we weigh the errors of the target class accordingly to allow the network to learn the underlying data distribution instead of resorting to predict the dominant zero-class. The benefit to this approach is that the network can learn subject-specific features from the PPG. However, the main drawback is that the network will need to be trained for each new user. In a commercial setting this would be very costly, since data would have to be collected during an initial enrollment period, networks would have to be trained, and weights downloaded back to the device. The training cost could be minimized by fine-tuning [9]. However, this formulation is natural, because it would not be feasible to implement this problem as multi-class classification since training examples of every possible customer would not be available during the initial training period. Furthermore, it is not practical to have a class for each subject since that would lead to an unbounded number of classes.

Along these lines, we believe it's quintessential to provide a brief description on [7], since it is one of the most recent investigations using wrist-worn, 'green' PPG under motion, which proposes a two-step process. The first step clusters the subjects based on 11 extracted features. These clusters are formed based on attributes such as gender or physical condition, but the authors also mention different forms of motion would fall under different clusters. This means that a single person could end up in more than one cluster depending if the given PPG was taken at rest, walking, biking, or any other ambulatory motion. Once these clusters have been formed, they implement a deep fully connected neural network to classify the subject within the cluster. Using features is susceptible to bias in generating the features and limits the

ability of the network to learn fully from the data. In this approach, a model is required for each group which still proposes significant overhead. Additionally, during this two-step approach, if the subject is not correctly filtered into the correct segment initially, there is no possibility to get the correct identification. Hence, to overcome these shortcomings, we propose a fully data-driven personalized deep learning approach.

III. PROPOSED FRAMEWORK

An overview of our framework is presented in Fig. 1. The PPG data samples are pre-processed with a band-pass 4th order Butterworth filter having the cut-off frequencies 0.1 – 18 Hz, which primarily restricts the high frequency noise component and drifts from the signal of interest. The signal is further normalized to zero mean and unit variance.

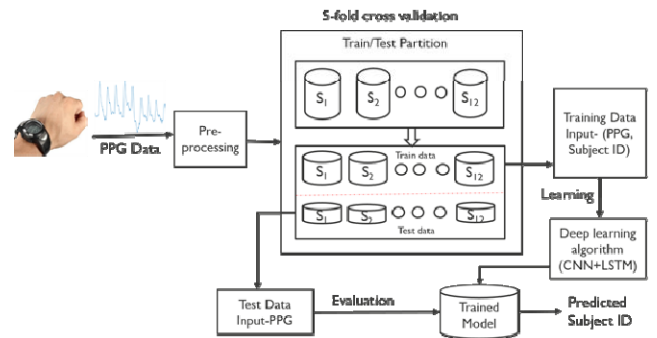


Fig. 1. Overview of the proposed methodology for biometric identification

A. DNN: CNN + LSTM

DNN allows feature extraction directly from in-domain data, thus enabling the learning of task-adapted feature representations [10]. The taxonomy of deep neural models mainly includes Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Stacked Auto-encoders. CNNs are characterized by an initial layer of convolutional filters (a set of weights which are slid over the input), followed by non-linearity (activation function – rectified linear units), sub-sampling (pooling), and a fully connected layer which realizes the classification. LeNet [11] was one of the very first CNNs which helped propel the field of Deep Learning. CNNs often consist of a stacked filter, activation, and pooling layers to enable the network to integrate the information from the different filters and various levels of abstraction. The stacking of multiple convolutional layers helps to achieve automatic feature extraction, where denser layers capture more complex or differentiating features.

Analyzing time series data often necessitates inferring the sequential/time-variant information. This is where RNNs prove to be an effective choice since they are able to incorporate contextual information from past inputs, having the advantage of being robust to localized distortions in the input sequence along the time. One key bottleneck in deep CNN structures (with an increased network capability) is the long chain problem (vanishing gradient) wherein information from previous computations rapidly attenuates as it progresses through the data flow [12]. A similar problem exists when

RNNs are applied to long sequential data, all the time steps have the same weight and, consequently, the contribution of an input in the hidden state is subjected to exponential decay. This is where a variant of RNN, LSTM comes in handy, in which the scalar valued hidden neuron is replaced with the LSTM memory block [12]. The LSTM memory block is inspired by a computer memory cell where context-dependent input, output, and forget gates control what is written to, read from, and kept in the cell in each time-step. In this way, it is easier for the network to store a given input over many time steps, in effect helping LSTM layers to capture temporal properties.

B. BiometricNET Architecture

In order to realize the data driven approach we allow the network to learn the discriminatory features, accomplished by using a two layer 1-D CNN. The output from the 1-D CNN is then fed into two LSTM layers and finally, the LSTM output is passed through a dense layer with SoftMax function for classification, constituting *BiometricNET*. The 1-D CNN can be thought of as a feature extractor. The input is convolved with the filters to generate points in the temporal-feature domain. Corresponding layers use these features to convolve with additional filters to generate the final features from the time-series PPG input. A drawback of using the CNN is that the generated features are not completely phase invariant. Depending on the time of occurrence of the heart beat relative to the beginning of the sample, the relevant features will be slightly shifted. Hence our use of LSTM which is primarily instrumental in capturing the temporal dependency in the sequence of historical local trends of the underlying cardiac activity inherent in the PPG signals further helps to recover from the phase offset.

C. Implementation Details

The PPG identification network was trained on an Nvidia Tesla K80 GPU and is modeled in Keras 2.0.4 version [13] configured to use theano [14] version 0.0.9 as the backend. Each CNN layer consists of a 1-D CNN operation with Scaled Exponential Linear Unit (SELU) activation [15], whereas each LSTM layer used hyperbolic tangent (tanh) function. The max-pooling layers used a pool size of 4, and a dropout layer with rate 0.1. Root Mean Square Propagation (RMSProp) is used as the optimizer with the default hyperparameters which is

recommended for training recurrent networks [16]. As described earlier the class loss is weighted to offset the class imbalance. Training batch size is set to 25 to balance training time and sensitivity to individual inputs. Hyperparameters such as filter length and number of filters, layers, and LSTM units were optimized using a heuristic grid search method, details of which are presented in the following section. Fig. 2 illustrates the proposed *BiometricNET* topology for this exploration.

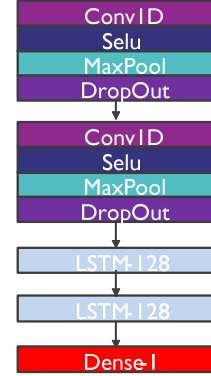


Fig. 2. Network topology used in the proposed *BiometricNET* using 2 CNN and LSTM layers in conjunction with a dense layer.

IV. RESULTS

The details of the hyperparameter grid search are presented in Table I. The performance of the model in conjunction with a given set of hyperparameters has been evaluated with the help of standard metrics, described briefly. The results in Table I represent the average outcomes of the various metrics as calculated on the 12 subjects, wherein C_1 and C_2 represents each CNN layer; sF and nF represents the filter size and number of filters respectively.

- a) *Accuracy*- ratio of correctly classified observation to the total observations.
- b) *Precision*- ratio of correctly predicted positive observations to the total predicted positive observations.
- c) *Recall*- ratio of correctly predicted positive observations to the all observations in actual class.
- d) *F1 Score*- weighted average of precision and recall.

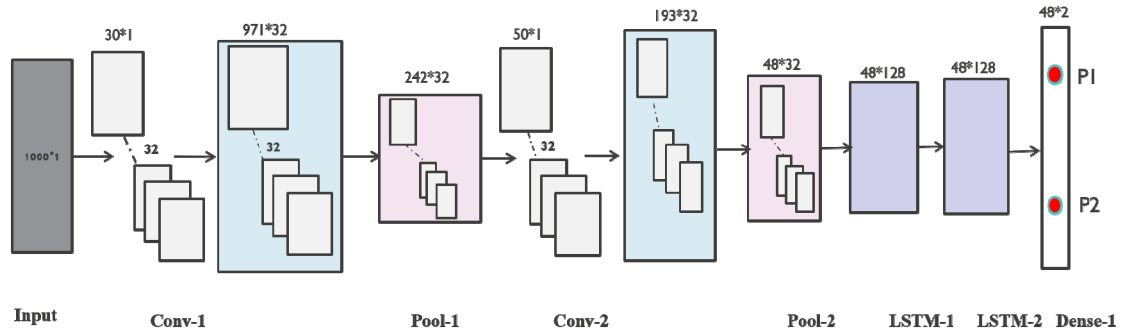


Fig. 3. *BiometricNET* architecture for identification, illustrating an input dimensionality of 1000 in conjunction with a filter size (sF) of 30 and 50 in layer1 and 2 (C_1 , C_2) respectively, each having a filter length (nF) of 32 and 128 LSTM units. P1 and P2 represent the probability of classified subject ID's. The remaining numbers are calculated based on a convolution operation with the input data.

TABLE I. PERFORMANCE ANALYSIS OF NETWORK CONFIGURATIONS.

Network Type	Network Configuration				A	F1	R	P
	sF		nF					
	C ₁	C ₂	C ₁	C ₂				
1	15	20	15	15	0.95	0.77	0.75	0.82
	15	30	15	15	0.95	0.80	0.77	0.85
	15	40	15	15	0.95	0.80	0.77	0.85
	15	50	15	15	0.96	0.82	0.81	0.85
2	30	20	15	15	0.95	0.81	0.78	0.86
	30	30	15	15	0.95	0.82	0.79	0.86
	30	40	15	15	0.96	0.82	0.80	0.89
	30	50	15	15	0.96	0.83	0.80	0.89
3	30	20	32	15	0.95	0.81	0.78	0.86
	30	30	32	32	0.96	0.84	0.81	0.88
	30	40	32	32	0.96	0.84	0.81	0.89
	30	50	32	32	0.96	0.86	0.84	0.89

A=Accuracy, F1=F1-Score, R=Recall, P=Precision

It can be observed from Table I that in all three network types, configuration with filter size (sF) 50 in second layer (C_2) performs comparatively better (highlighted in bold). Furthermore, among the selected best configuration in three network types, the network with a filter length (nF) of 32 and sF of 30 and 50 (marked in bold), achieved the best performance demonstrating an improvement of approximately 2 and 4% in f1 score and recall value respectively compared to the other two network types. Hence, we can conclude that the configuration 30-50-32-32 provides the best performance for our investigated problem, demonstrating an average accuracy, f1 score, recall, and precision of 0.96, 0.86, 0.84 and 0.89 respectively. Increasing the sF and nF values, beyond this did not improve the overall performance. It is important to mention that a similar exhaustive exploration was also done with the LSTM size and the selected 128 units for both layers provided the best performance (details have not been shown for sake of brevity) in conjunction with the demonstrated CNN architecture. The dense layer used at the end in conjunction with a softmax function has 2 neurons. Discriminatory weights pertaining to the first CNN layer for extraction of two example subjects (1 and 5) have been shown in Fig. 4.

The chosen network model has been illustrated in Fig. 3 and the evaluation results over 12 subjects have been presented in Table II. It can be observed that subject 4 and subject 9 have the best performance in terms of average accuracy (98%) and precision (98%). For real-time execution on an embedded device, driven by application requirements, a complexity analysis of the proposed architecture was performed. A single evaluation of the four layer network requires approximately 452K MACs (multiply-and-accumulate). Current trends in architecture development hold promise and make this number achievable [17].

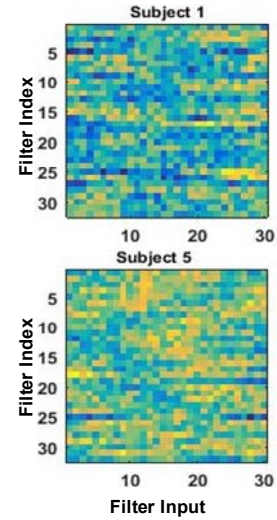


Fig. 4. An illustrative example of the first CNN layer filter weights for networks trained for subjects 1 and 5

TABLE II. PERFORMANCE ANALYSIS OF THE SELECTED NETWORK ($sF = 30, 50$; $nF = 32, 32$) FOR BIOMETRIC IDENTIFICATION

Subject	Average (5-CV) Accuracy	F1 Score	Recall	Precision
S1	0.95	0.75	0.68	0.86
S2	0.96	0.83	0.82	0.83
S3	0.94	0.73	0.76	0.70
S4	0.98	0.92	0.87	0.98
S5	0.96	0.83	0.90	0.79
S6	0.96	0.88	0.83	0.93
S7	0.94	0.73	0.67	0.82
S8	0.98	0.94	0.94	0.94
S9	0.98	0.94	0.91	0.98
S10	0.98	0.95	0.95	0.97
S11	0.97	0.90	0.88	0.93
S12	0.98	0.94	0.91	0.97
Mean	0.96	0.86	0.84	0.89

V. CONCLUSION

This paper presents a first of its kind exploration where we move away from the need for extracting hand crafted features for biometric identification and report on a personalized data-driven approach using deep learning on wearable PPG signals collected in ambulatory situation. Our *BiometricNET* topology, using two layers each of CNN and LSTM yields best accuracy of 98% and an average accuracy of 96% on all 12 subjects. In view of the present work, future explorations would focus on evaluating the network on other databases having wrist PPG signals acquired amidst motion. Future research will focus on energy-efficient execution of the algorithm on wearable devices in real-time on a microcontroller or android for mobile platform or hardware solutions (ASIC, SoC) using the schemes proposed in [17] [18].

REFERENCES

- [1] Bonissi, R. Donida Labati, L. Perico, R. Sassi, F. Scotti, and L. Sparagino, "A preliminary study on continuous authentication methods for photoplethysmographic biometrics", in Proc. of the 2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS 2013).
- [2] A. Lourenço, H. Silva, A. Fred, "Unveiling the biometric potential of finger-based ecg signals", Computational intelligence and neuroscience, vol. 2011, pp. 5, 2011
- [3] Y. Gu, Y. Zhang, Y. Zhang, "A novel biometric approach in human verification by photoplethysmographic signals", 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, pp. 13-14, 2003.
- [4] Y. Gu, Y. Zhang, "Photoplethysmographic authentication through fuzzy logic", IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, pp. 136-137, 2003.
- [5] P. Spachos, J. Gao, D. Hatzinakos, "Feasibility study of photoplethysmographic signals for biometric identification", 17th IEEE International Conference on Digital Signal Processing (DSP), pp. 1-5, 2011.
- [6] Kavsaoglu, A. Reşit, Kemal Polat, and M. Recep Bozkurt, "A novel feature ranking algorithm for biometric recognition with PPG signals." Computers in biology and medicine 49 (2014): 1-14.
- [7] Jindal, Vasu, et al. "An adaptive deep learning approach for PPG-based identification." Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the IEEE, 2016.
- [8] Zhang, Zhilin, Zhouyue Pi, and Benyuan Liu. "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise." IEEE Transactions on Biomedical Engineering 62.2 (2015): 522-531.
- [9] Lalor, John P., Hao Wu, and Hong Yu. "CIFT: Crowd-Informed Fine-Tuning to Improve Machine Learning Ability."
- [10] Bengio, Yoshua. "Learning deep architectures for AI." Foundations and trends® in Machine Learning 2.1 (2009): 1-127.
- [11] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.
- [12] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.
- [13] <https://keras.io/> [online]
- [14] <http://deeplearning.net/software/theano/> [online]
- [15] Klambauer, Günter, et al. "Self-Normalizing Neural Networks." arXiv preprint arXiv:1706.02515 (2017).
- [16] <http://climin.readthedocs.io/en/latest/rmsprop.html> [online]
- [17] Han, Song, Huizi Mao, and William J. Dally. "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding." arXiv preprint arXiv:1510.00149 (2015).
- [18] M. Panwar, J. Padmini, N. Venkatasubrahmanian, A. Acharyya, and D. Biswas, "Modified Distributed Arithmetic Based Low Complexity CNN Architecture Design Methodology", 23rd IEEE European Conference on Circuit Theory and Design (ECCTD), Catania, Italy, September 4-6, 2017.