FUSION-ID: A PHOTOPLETHYSMOGRAPHY AND MOTION SENSOR FUSION BIOMETRIC AUTHENTICATOR WITH FEW-SHOT ON-BOARDING

Harshat Kumar[*], *Hojjat Seyed Mousavi* [†], *Behrooz Shahsavari* [†]

Apple Inc.

ABSTRACT

The abundance of wrist-worn heart rate measuring devices enables long term cardiovascular monitoring through photoplethysmography (PPG). Such signals contain unique identifiable information that can help in biometric authentication. In this work, we propose Fusion-ID, which use wrist-worn PPG sensors fused with motion sensor data as a way to do bio authentication on wrist worn devices. We conducted a user study using a PPG and motion sensor enabled wrist-worn device and collected data from 247 users. We then propose a novel sensor fusion deep Siamese network architecture for feature embedding. Specifically, Fusion-ID fuses information from multiple channels of PPG readings with information from motion sensors to authenticate a user. Our architecture only needs a few seconds of sample data (shots) from new users, and it is the first PPG-based bio-authentication method that is capable of adapting to new users without requiring on-device training or fine-tuning of the model. Our evaluations confirm the effectiveness of proposed siamese network with sensor fusion with an average accuracy of 95% and 12% false rejection rate at the 1% false acceptance rate operating point.

***Index Terms*—** PPG, IMU, Siamese, Biometric, Authentication, Few-shot

1. INTRODUCTION

Knowledge-based and passcode-driven authentication methods have some inherent risks, such as forgetting passwords or shoulder surfing attacks. Therefore, the use of biometrics as an authentication mean has recently gained significant interest [1]. To this end, sensors available on smartwatches provide meaningful and unique information, which have been shown to be helpful to authenticate users by their gait [2], wrist movement [3], and behavior [4].

Many smartwatches are equipped with PhotoPlethysmoGraphy (PPG) sensors, which measure the change in blood volume from heart activity. PPG signals are non-invasive, easy to measure, accessible from diverse body positions, and hold distinctiveness, which makes them a candidate [5] for passive, *gesture-less* authentication. Although it has been shown that PPG signals can be used to differentiate users, a discussion of feasibility to extend these systems to new users is lacking. Plus PPG signals are prone to movements and can exhibit random behavior. Smartwatches are also usually equipped with Inertial Measurement Unit (IMU) sensors, which capture motion through accelerometers and gyroscopes. IMUs have also demonstrated success in authentication through the uniqueness of an individual's gait [6]. In this work, we combine PPG and IMU based authentication by fusing the two to reduce the inherent randomness of PPG based authentication.

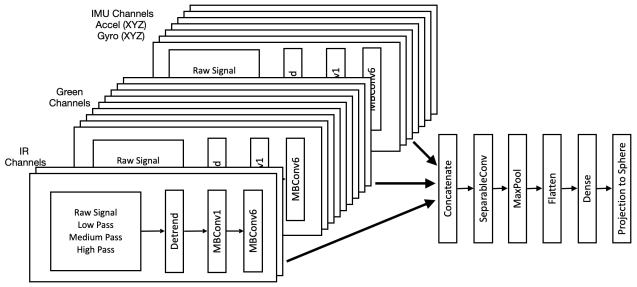


Fig. 1. Feature Extractor architecture consisting of PPG (IR and Green channels) and motion sensor (IMU channels) data.

The main contributions of the proposed Fusion-ID are as follows. First, we conduct a user study with 247 users of varying demographics. Second, we introduce a deep sensor fusion neural network combining PPG and IMU channels. Finally, we consider a Siamese learning structure that extends to new users with only a few examples, or shots, *without* the need to change the model. To the best of our knowledge, we are the first to consider this PPG-based few-shot on-boarding of new users.

2. RELATED WORK

In the past two decades, bio-authentication has been an active area of research, notably starting with authentication with ElectroCardioGram (ECG) signals. Since their proof of concept, ECG signals have been proven to be effective first through fiducial features, such as amplitude, on-set time, duration, slope [7], then through non-fiducial methods such as, discrete wavelet transform and principal component analysis [8], and finally through deep neural network (DNN) architectures such as recurrent neural networks [9] and siamese networks [10].

PPG signals have also shown a similar trend, moving away from fiducial authentication towards authentication with DNNs. Fiducial methods include the use of fuzzy logic [11], second derivatives [12], and sum of gaussians [13]. Moving away from fiducial methods, linear discriminant analysis (LDA) has been used for authentication, showing success on users with varying emotions and physical activities [14].

PPG signals have also been considered with other signals to improve authentication. For example, they have been used with ECG as part of a two factor authentication [15]. Similarly, PPG sensors

[*]Dept of Electrical and Systems Engineering, University of Pennsylvania (Work was done during an internship at Apple Inc.)

[†]Apple Inc.

were combined with IMU sensors to determine which specific statistical features best separated users for the purpose of authentication, albeit on custom built hardware [16].

Recently, there has been an effort to use DNNs to improve the performance of PPG based authentication. Notably, there are two schools of thought. The first focuses on data driven authentication without any preprocessing [17]. They proposed two convolution neural network (CNN) and two long short term memory (LSTM) layers which showed effectiveness in heart rate estimation [18]. On the other hand, a multi modal approach which uses cubic interpolation, dynamic warping, Fourier and inverse Fourier transforms to stretch the signal, that is scale and crop the target wave, has also been considered [19]. These methods are used to stretch the signal to provide a time stable input to a DNN, and showed high accuracy for users across multiple sessions [20]. Additionally, PPG signals have been considered for gesture-based authentication [21] as well as for spoof detection [22].

3. PROPOSED FUSION-ID

We describe our Fusion-ID approach to the PPG and IMU based authentication problem in two parts, offline training of the feature extractor, and on-boarding of new users. During offline training, we train a siamese network to obtain an informative feature embedding of the PPG and IMU signals. During new user on-boarding and evaluation, a support set is generated per user, and a query is authenticated if the distance score between the new embedding and the support set is less than a specific threshold. Before we describe the specifics of these two tasks, we introduce the following notation.

Let $\mathcal{X}$ be the offline dataset used to train the model embedding with m users and k_i samples per user $i = 1, \ldots, m$. Similarly, let $\mathcal{X}^*$ be the set of m^* new users each with k_i^* samples per user. We denote a sample of $\mathcal{X}$ by $x_i^j(^*)$ where i denotes the user and j denotes the sample index of that user, and the $*$ denotes whether or not the sample comes from the offline datasets or new users.

3.1. Offline Training

Let f_θ be a feature embedding parameterized by θ. The goal for offline training is to find θ^*, an embedding that minimizes the distances between samples from the same user and maximizes the distances from different users. To achieve this goal, we consider training a siamese network with triplet loss.

The triplet loss is defined on a tuple consisting of an anchor, a similar, and a dissimilar sample (x_i^a, x_i^b, x_j^c, respectively) with some margin γ by

$$loss(x_i^a, x_i^b, x_j^c) = \max\{\|f_\theta(x_i^a) - f_\theta(x_i^b)\| - \|f_\theta(x_i^a) - f_\theta(x_j^c)\| + \gamma, 0\}.$$

Our objective is to minimize the loss. If the distance between the anchor and dissimilar samples is larger than the distance to the anchor and similar samples by γ, the loss is zero. The positive margin γ is a design parameter which may be tuned during training.

Preprocessing: In this study, we consider signals of five seconds or ten seconds as inputs to the Fusion-ID network for passive authentication. The raw input of the model is a signal consisting of 16 channels. These include two channels which come from the infrared (IR) signals sampled at 64Hz and eight green channels sampled at 256Hz. We additionally include IMU signals, which consists of six channels sampled at 100Hz (three for accelerometer and three for gyroscope). We choose to include the IMU channels to capture motion artifacts, which help further distinguish between users. Each raw input is preprocessed by Butterworth filters. For the infrared and IMU, we use filters with different pass band spectrums of (0.25-8 Hz, 8-32 Hz, and $>$ 32 Hz). For the green channels, we use an additional filter to have filters of (0.25-2 Hz, 2-8 Hz, 8-32 Hz, and $>$ 32 Hz).

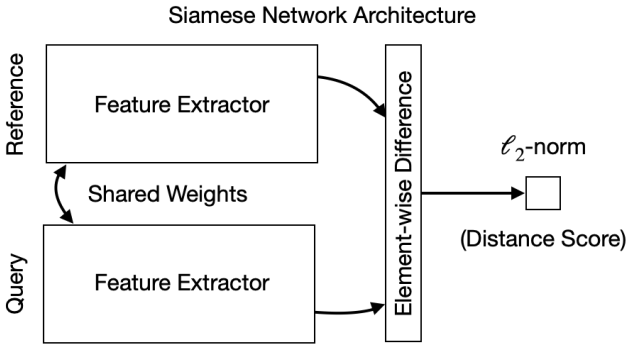


Fig. 2. Siamese network consists of two feature extractor with tied weights. After the feature extractor, the point-wise difference is taken, and the l_2 norm gives the distance score.

Feature embedding: The feature embedding architecture is shown in Fig 1. Using a similar architecture as EfficientNet [23], for each input channel, we employ two inverted residual blocks [24]. After this per channel feature extraction step, we concatenate the output of all the channels and apply a separable convolution layer, followed by max pooling and a flatten layer. Next, we add a fully connected dense layer followed by a projection on a unit hypersphere. The projection step is important to avoid the trivial solution of siamese networks with triplet loss (see Remark 1).

Siamese architecture: Given the feature embedding, we can now describe the siamese architecture used during both training and evaluation which is shown in Fig 2. The siamese network takes two inputs and computes their feature embedding. The two feature extractors have shared weights. Upon obtaining the feature embedding, the Siamese network computes the ℓ_2-norm of their element-wise difference. During offline training, this distance is used to compute the loss.

Remark 1 : The projection step for the feature extraction network is important as the triplet loss is otherwise prone to converge to a trivial solution where the embedding maps all signals to the same point. We include this remark as the discussion was lacking in the literature.

3.2. Few shot on-boarding of new users

Given the trained feature embedding f_{θ^*}, the objective of on-boarding new users is to correctly authenticate users unseen by the model during training. This objective is achieved by selecting a support set of N samples and a positive threshold $\tau > 0$. In particular, for each new user in $\mathcal{X}^*$, we select $\mathcal{S}_i := \{x_i^1, x_i^2, \ldots, x_i^N\}$ to be the support set. The remaining $\{x_i^{N+1}, \ldots, x_i^{k_i}\}$ samples are used as queries, denoted as s, for evaluation, representing the instances after on-boarding, where the user wishes to authenticate. Given the support set, we authenticate the user when the minimum

distance of the support set to the query is less than the threshold τ. The minimum distance is described formally as

$$d_{\text{min}}(s, \mathcal{S}_i) := \min_a \{\|f_{\theta^*}(s_i^a) - f_{\theta^*}(s)\|\}_{a=1,\ldots,N}.$$

The authentication scheme therefore becomes

$$s \in \mathcal{S}_i, \text{ when } d_{\text{min}}(s, \mathcal{S}_i) \leq \tau$$
$$s \notin \mathcal{S}_i, \text{ when } d_{\text{min}}(s, \mathcal{S}_i) > \tau.$$

It has been demonstrated empirically that PPG signals change over time [20]. Accordingly, we consider the minimum distance to the support set due to the high variance we expect to see over time.

For the purpose of this study, we consider five, ten, and fifteen shots to onboard a new user. The model is only required to save these support samples for comparison against new queries. To that end, our few shot learning paradigm easily extends to new users *without the need for any fine tuning or model training on device*.

4. RESULTS AND DISCUSSION

In this section, we evaluate the performance of the trained feature embedding on the evaluation dataset of new users. We emphasize that the model has not seen any data from these users during training, and on-boarding is only recording N samples to use as support, as described in the previous section.

Dataset: We collected data from 247 users. We continuously collect samples over the course of an hour to account for variation in time. We segment the data into five and ten second input signals. We partitioned the data into 194 users for training the offline feature extractor and the remaining 24 as a validation set. We evaluate the model on the remaining 29 users, which will act as the novel users. Ranging from 21 to 63, the average age of the participants is 33. The data collected from the female, male, and non-binary participants came included a variety of body postures (standing, sitting, laying down, and walking) and other routine tasks (typing, talking on the phone, washing dishes, and using mobile phone). All participants wore the watch on their left hand.

4.1. Evaluation Metrics

To on-board new users, we select N samples for the support of that user and the remaining samples for evaluation. The N on-boarding samples represent the number of shots required to onboard the user. To evaluate our method, we compute the false acceptance rate (FAR), false rejection rate (FRR), and the equal error rate (EER), which are defined on the threshold τ as

$$\text{FAR}(\tau) = \frac{\text{FalseAcceptance}(\tau)}{n_{\text{Imposter}}},$$

$$\text{FRR}(\tau) = \frac{\text{FalseRejection}(\tau)}{n_{\text{Genuine}}},$$

$$\text{EER} = \text{FAR}(\tau_{\text{equal}}) \equiv \text{FRR}(\tau_{\text{equal}}),$$

where

$$\tau_{\text{equal}} = \arg\min_\tau \|\text{FAR}(\tau) - \text{FRR}(\tau)\|.$$

Additionally, we show the receiver operating characteristic (ROC) curves (see Fig 5) to compare models agnostic to the threshold τ. The ROC curve shows the FRR corresponding to each FAR at various threshold settings. Finally, because we consider authentication, we also evaluate the FRR at the operating thresholds where the FAR is equal to 1% and 5% (see Fig 3 (c-d)). Evaluating the models at these operating points is realistic to consider for the authentication problem because the cost of falsely authenticating an imposter is higher than falsely rejecting the user. The latter case can actually improve the authentication model, as we will discuss in our future works Section 5. Note also that for all of the metrics, every user was evaluated on the same threshold τ_{equal}. By personalizing thresholds at the user level, we would expect to see improved performance.

4.2. Results

In Fig 3, we compare the performance of Fusion-ID with varying support samples and input channels for the case of a ten second input. We find that increasing the number of input channels significantly improves the performance of the model on every performance metric. This shows the effectiveness of Fusion-ID in combining the input channels in a meaningful way. Increasing the number of support samples also improves performance. Saving a lot of samples is reasonable to consider for wrist-worn heart rate measuring devices, as only the low dimensional embedding needs to be saved to the device's memory.

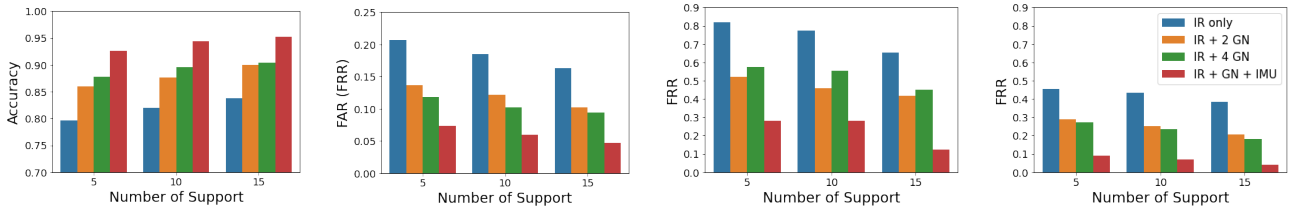
A summary of the performance of the five second input compared to the ten second input with $N = 15$ support samples can be found in Table 1. Highlighted in bold, Fusion-ID with with both PPG and IMU channels and ten second inputs achieves the best accuracy of 95%, and a 12% FRR at the 1% FAR operating point. Adding green channels to the IR only model improves both the ten second and five second input models significantly (6% accuracy increase for ten and 9% for five seconds). The per user accuracies for Fusion-ID with five and ten second inputs are shown in Fig 4. With the exception of very few users, the ten second input significantly outperforms the five second input. Longer input signals are reasonable to consider, as the authentication should occur before the user's first interaction with the device as well as the extension to continuous authentication [22].

We emphasize that this performance is on users that the model did not use to create the feature embedding, a paradigm not considered by related PPG based authentication works.

4.3. Baseline comparison

As discussed earlier, previous related work does not extend to new users without the need to fine tune or change the model on device [18, 20]. We therefore compare our learned feature extractor against classical dimension reduction techniques. Specifically, we consider independent component analysis (ICA) and linear discriminant analysis (LDA) where the channel signals are concatenated to create one high dimensional vector [25, 26]. Fitting the ICA and LDA transforms on the offline data of 194 users, we apply the transform on the new user samples to generate features in $\mathbb{R}^{100}$. We use the same distance score with ℓ_2-norm to predict each query.

Figure 5 shows the performance of Fusion-ID against ICA and LDA. To highlight the differences between them, we plot the ROC curve with the FAR on logarithmic scale. Our method outperforms both baselines by a significant margin. Most interestingly, our method performs well on the operation point of low FAR ($< 10^{-2}$). This is important while considering authentication methods as the cost of falsely authenticating an imposter is much higher than falsely



(a) Accuracy (b) EER (FRR = FAR) (c) FRR at (FAR = 0.01) (d) FRR at (FAR = 0.05)

Fig. 3. Evaluation metrics with varying input channels (see legend in (d)) and size of support set. Size of input is ten seconds.

	10 Seconds			5 Seconds		
Input Type	Accuracy	FRR at 0.01 FAR	FRR at 0.05 FAR	Accuracy	FRR at 0.01 FAR	FRR at 0.05 FAR
IR Only	0.84	0.65	0.38	0.75	0.75	0.56
IR + 2 GN	0.90	0.42	0.21	0.84	0.49	0.31
IR + 4 GN	0.90	0.45	0.18	0.87	0.57	0.26
IR + GN + IMU	**0.95**	**0.12**	**0.04**	0.87	0.41	0.24

Table 1. Comparison of five and ten second input signals with varying number of input channels for $N = 15$ support samples.

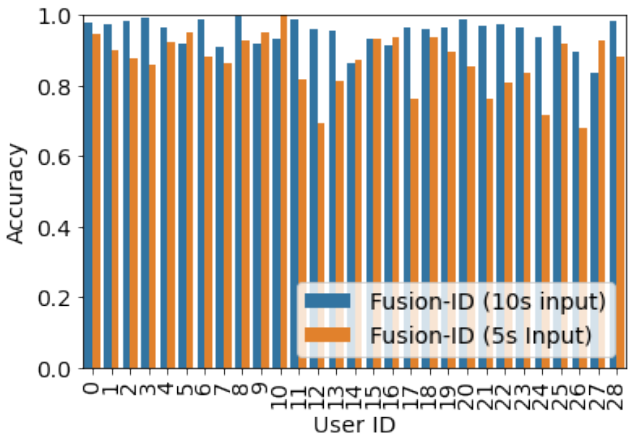


Fig. 4. Accuracies per user with the 10 second input signal (blue) and the 5 second input signal (orange). 10 second input enjoys 8% average accuracy gain over 5 second inputs (95% compared to 87%). Number of support samples $N = 15$ with all input channels (IR, Green, and IMU)

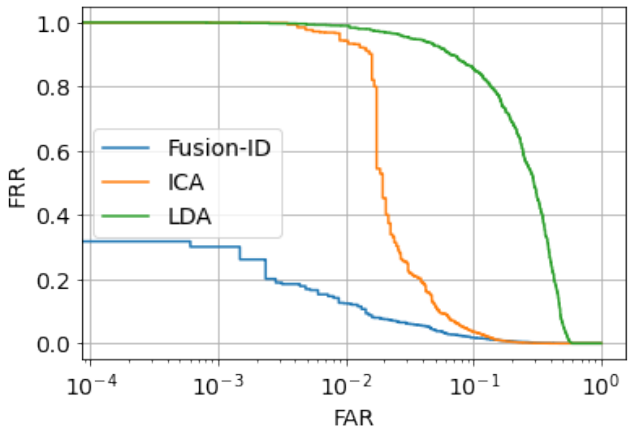


Fig. 5. Receiver operating characteristic (ROC) curve comparing Fusion-ID against the baselines of ICA and LDA. The FPR is shown in log scale.

rejecting a valid user. It is precisely in this low FAR operation region where our method outperforms the others by the largest margin.

5. CONCLUSION AND FUTURE WORK

In this work, we considered PPG-based authentication fused with motion data. We trained the feature extractor using a siamese model with triplet loss and evaluated on a new set of users not seen by the model during training. Using only a few on-boarding shots, we showed that our method achieves an average accuracy of 95%, outperforming traditional feature extraction methods. Fusion-ID works particularly well on low FAR operating points, which are essential for ensuring that imposters are not falsely authenticated.

There are a number of possible directions for future work. For example, a long term feasibility study should be considered, where the multiple sessions take place over the course of multiple days. In our work, we considered data from different sessions of a user collected over a course of an hour; however, it is possible that there is more variation that occurs day to day. We believe increasing the support set will reduce adverse effects form long term variability. Adaptively or actively learning support set can be used to reduce such effects. In the case of a false rejection, the user will need to authenticate themselves using a knowledge factor based method. This provides the model with a labeled shot for free, which the model can then use to either add to or update their support set. Finally, this method may be used to identify imposters who know the user's knowledge based factors. Naively, this may be achieved by setting an upper threshold, such that if the minimum distance is greater than that upper threshold, the device may reject the user.

6. REFERENCES

[1] Shuqi Liu, Wei Shao, Tan Li, Weitao Xu, and Linqi Song, "Recent advances in biometrics-based user authentication for wearable devices: A contemporary survey," *Digital Signal Processing*, p. 103120, 2021.

[2] Weitao Xu, Chitra Javali, Girish Revadigar, Chengwen Luo, Neil Bergmann, and Wen Hu, "Gait-key: A gait-based shared secret key generation protocol for wearable devices," *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 1, pp. 1–27, 2017.

[3] Alona Levy, Ben Nassi, Yuval Elovici, and Erez Shmueli, "Handwritten signature verification using wrist-worn devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–26, 2018.

[4] Ingo Deutschmann, Peder Nordström, and Linus Nilsson, "Continuous authentication using behavioral biometrics," *IT Professional*, vol. 15, no. 4, pp. 12–15, 2013.

[5] Mohamed Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current cardiology reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[6] Omid Dehzangi, Mojtaba Taherisadr, and Raghvendar ChangalVala, "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors*, vol. 17, no. 12, pp. 2735, 2017.

[7] Lena Biel, Ola Pettersson, Lennart Philipson, and Peter Wide, "ECG analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[8] Se Young Chun, "Single pulse ECG-based small scale user authentication using guided filtering," in *2016 international conference on biometrics (ICB)*. IEEE, 2016, pp. 1–7.

[9] Ronald Salloum and C-C Jay Kuo, "ECG-based biometrics using recurrent neural networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2062–2066.

[10] Nabil Ibtehaz, Muhammad EH Chowdhury, Amith Khandakar, Serkan Kiranyaz, M Sohel Rahman, Anas Tahir, Yazan Qiblawey, and Tawsifur Rahman, "Edith: ECG biometrics aided by deep learning for reliable individual authentication," *arXiv preprint arXiv:2102.08026*, 2021.

[11] YY Gu and YT Zhang, "Photoplethysmographic authentication through fuzzy logic," in *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.* IEEE, 2003, pp. 136–137.

[12] Nur Azua Liyana Jaafar, Khairul Azami Sidek, and Siti Nurfarah Ain Mohd Azam, "Acceleration plethysmogram based biometric identification," in *2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*. IEEE, 2015, pp. 16–21.

[13] Abhijit Sarkar, A Lynn Abbott, and Zachary Doerzaph, "Biometric authentication using photoplethysmography signals," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2016, pp. 1–7.

[14] Umang Yadav, Sherif N Abbas, and Dimitrios Hatzinakos, "Evaluation of PPG biometrics for authentication in different states," in *2018 International Conference on Biometrics (ICB)*. IEEE, 2018, pp. 277–282.

[15] Lucas Bastos, Thais Tavares, Denis Rosário, Eduardo Cerqueira, Aldri Santos, and Michele Nogueira, "Double authentication model based on PPG and ECG signals," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2020, pp. 601–606.

[16] Guannan Wu, Jian Wang, Yongrong Zhang, and Shuai Jiang, "A continuous identity authentication scheme based on physiological and behavioral characteristics," *Sensors*, vol. 18, no. 1, pp. 179, 2018.

[17] Luke Everson, Dwaipayan Biswas, Madhuri Panwar, Dimitrios Rodopoulos, Amit Acharyya, Chris H Kim, Chris Van Hoof, Mario Konijnenburg, and Nick Van Helleputte, "Biometricnet: Deep learning based biometric identification using wrist-worn PPG," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5.

[18] Dwaipayan Biswas et al., "CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment," .

[19] Dae Yon Hwang, Bilal Taha, Da Saem Lee, and Dimitrios Hatzinakos, "Evaluation of the time stability and uniqueness in PPG-based biometric system," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 116–130, 2020.

[20] Dae Yon Hwang, Bilal Taha, and Dimitrios Hatzinakos, "Variation-stable fusion for PPG-based biometric system," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8042–8046.

[21] Yu Ling, Xiang Chen, Yuwen Ruan, Xu Zhang, and Xun Chen, "Comparative study of gesture recognition based on accelerometer and photoplethysmography sensor for gesture interactions in wearable devices," *IEEE Sensors Journal*, 2021.

[22] Abbas Acar, Hidayet Aksu, A Selcuk Uluagac, and Kemal Akkaya, "A usable and robust continuous authentication framework using wearables," *IEEE Transactions on Mobile Computing*, vol. 20, no. 6, pp. 2140–2153, 2020.

[23] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.

[24] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[25] Chen He and Jane Wang, "An independent component analysis (ICA) based approach for EEG person authentication," in *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2009, pp. 1–4.

[26] Prashant Kumar Jain, Shailja Shukla, and SS Thakur, "Fuzzy fusion of PCA, ICA and ILDA face algorithms for enhanced user authentication," *Journal of Engineering Science and Technology*, vol. 12, no. 9, pp. 2297–2314, 2017.