IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 October 2023, accepted 28 October 2023, date of publication 2 November 2023, date of current version 9 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329993

RESEARCH ARTICLE

PulseOblivion: An Effective Session-Based Continuous Authentication Scheme Using PPG Signals

HUSSEIN A. ALY[1] AND ROBERTO DI PIETRO[2], (Fellow, IEEE)
[1]College of Engineering, Department of Computer Science and Engineering, Qatar University, Doha, Qatar
[2]Resilient Computing and Cybersecurity Center (RC3), Computer, Electrical, and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Corresponding author: Hussein A. Aly (hussein.aly@qu.edu.qa)

This work was supported by the Open Access funding provided by the Qatar National Library.

ABSTRACT In this paper, we propose a novel session-based continuous authentication model using photoplethysmography (PPG). Unlike previous PPG-based authentication techniques that generate user signatures only during the initial interaction, our session-based approach tackles inter session PPG drifting by generating a user signature at the start of each session. Our model is composed by two modules: Firstly, heavy deep autoencoders (AE) are utilized for feature extraction and, secondly, a lightweight Local Outlier Factor (LOF) is employed for user authentication.Additionally, we introduce a continuous updating system for the LOF model, which automatically recovers from security breaches and can enhance authentication accuracy by more than 9%. Our experiments show that in a single-session scenario, our model achieves authentication accuracies of 93.5% and 91.8% on the CapnoBase and BIMDC benchmarking datasets, respectively, outperforming the state-of-the-art baseline model by 3.2% and 1.6% on both datasets, respectively. In multiple-session scenarios, our scheme attains an authentication accuracy of 95% when tested on the BioSec2 dataset, effectively mitigating inter-session PPG drifting and achieving an advantage of more than 8.5% in authentication accuracy over the state-of-the-art method. In terms of execution speed, our solution is seven times faster at runtime compared to competing state-of-the-art solutions.

INDEX TERMS Security, biometric authentication, continuous authentication, PPG, deep autoencoders.

I. INTRODUCTION

Authentication is a vital part of any security system, as it can be used to grant access to legitimate users only. Traditionally, many authentication systems rely on static authentication schemes, where the user is authenticated only once at the start of the session [1]. Despite being effective against various security threats such as brute force attacks, static authentication schemes give permanent access to the device and data, possibly leading to security breaches (e.g., hijacking the user session). A typical example is a bank employee that leaves their terminal unattended after having logged in. To cope with the vulnerabilities introduced above, continuous authentication schemes have been introduced. At its core, such a solution continuously checks user identity, locking down the system if the user identity could not be verified [2].

A simple solution for continuous authentication is to ask the user to input the password periodically; however, this approach would disrupt the user's workflow and reduce the system's overall usability. Another approach is to use one-time authentication and then use proximity sensing on user's wearable devices, such as a smartwatch or smartphone, to identify when the legitimate user is not in the proximity of system—unattended systems can easily be exploited by a physically close adversary. However, this latter approach is vulnerable if the adversary gain position of the user device, or the user unintentionally leaves it near the system. Alternatively, the authentication system could passively collect information about the user to be used in the authentication processes. The collected information can be behavioral characteristics, such as keystrokes and

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

mouse movement [3]. However, behavior characteristics require the user to maintain a continuous active interaction with the system (e.g., continuous mouse movement or keyboard usage), which would be inconvenient if the user is performing passive activities like reading an article or watching a video. These cited limitations could be overcome if the employed solution resorts to biometric signals, such as photoplethysmography (PPG) [4] or Electrocardiography (ECG) [5] or fusion of both [6]. Indeed, when the above introduced techniques are adopted, continuous authentication can be implemented, as the data to be inputted into the system are collected passively and does not require maintaining active user interaction with the system. Also, it is not possible for an adversary to deceive the system by utilizing even a stolen honest user's wearable device, since the data to support authentication are the biometric data of the owner of the device, the device itself not storing or producing any useful material to pass authentication. These features make biometric data perfect candidates for continuous authentication systems.

In the realm of biometric authentication, it is worth noticing the recent advancement on the Internet of Medical Things (IoMT) and biometric sensing. These advances resulted in the widespread diffusion of wearable devices that include biometric sensors such as fitness trackers and smartwatches that usually include sensors like PPG. In addition, ECG can provide more detailed information about the person compared to PPG and usually leads to a more accurate authentication system; however, ECG is inconvenient in real-life scenarios, since the sensors needed to implement the cited technique are more expensive and more difficult to wear and maintain [7]. Instead, PPG data is much easier to collect, and PPG sensors are much more cost-efficient than ECG sensors [7]. Also, PPG signals provide a solution to the vulnerability of traditional authentication systems against spoofing attacks, as it provides advantages such as being difficult to spoof or steal and live detection [8]. Furthermore, several works have demonstrated the feasibility of PPG-based authentication systems, as the distinctive differences in PPG signals between individuals can be utilized for unique identification [9], [10], [11].

While PPG-based authentication is still in its early stages [8], various spoofing attacks have been attempted against PPG-based authentication systems. One example is the use of stealthy recording of PPG signals from non-genuine measurement sites (also known as presentation attacks) to deceive PPG-based authentication systems [12]. Another approach is to remotely steal PPG signals from video recordings of the victim, known as remote PPG (rPPG), which can be used to breach the authentication system [13].

Therefore, in this article, we focused on developing a PPG-based authentication system that provides high usability in practical use cases and with high authentication accuracy. However, it should be noted that PPG based authentication is not an easy task. For instance, signals collected from wearable sensors are affected by the user's physical or mental state change. Thus, factors that can affect the cardiovascular system, such as diet, sleeping, physical exercise and emotional state affect the PPG signature of the user. This effect can lead to the increase in the False Rejection Rate (FRR), where honest users are denied access to the system, raising a need to provide an authentication scheme that is tolerant to fast change of PPG signal (or PPG time drifting).

Contributions

In this work, we provide several contributions. In particular:

- A session-based continuous PPG authentication scheme that eliminates the effect of inter-session PPG time drifting while providing higher authentication performance than traditional schemes and having similar buffering duration. As shown by our experiments, our proposed schemes provide an advantage of more than 8.5% in authentication accuracy compared to state-of-the-art schemes on the BioSec2 dataset with an inter-session gap of 17 days. Also, our scheme requires a registration period (buffer size) as short as 15 heartbeats (≈ 15 seconds assuming heart rate of 60 BPM) while providing over 91% F1 score when evaluated on the CapnoBase dataset.
- An efficient continuous authentication model using deep autoencoders (AE) and a lightweight Local Outlier Factor (LOF) model. AE requires training just once on a subset of users and can be applied to all users without retraining, increasing the model utility. LOF facilitates quick user registration and authentication in a short registration period (as low as 15 beats). Our solution surpasses traditional signal processing-based systems, achieving an F1 score of 95.9% 91.3% and 91.0% in the BioSec2, CapnoBase, and BIMDC datasets, respectively. Also, our proposed solution has a sevenfold faster execution time during authentication.
- A mechanism for continuously updating the authentication model that provides a self-healing property, enhancing the model's resilience to adversarial attacks and addressing intra-session PPG drifting, providing more than 9% increase in authentication accuracy.

Roadmap

This paper is organized as follows: in Section II we provide background information about authentication modes and PPG signals. Section III summarizes the related work on PPG authentication. Then, in Section IV, we discuss the proposed continuous authentication scheme and authentication model alongside the baseline model that was used to evaluate the Performance of our proposed solution. In Section V we analyse and discuss the findings of our experiment. Finally, Section VI summarizes our work and discusses future directions.

II. BACKGROUND

In this section, we discuss an overview of the PPG signal, followed by a survey of the various types of authentication

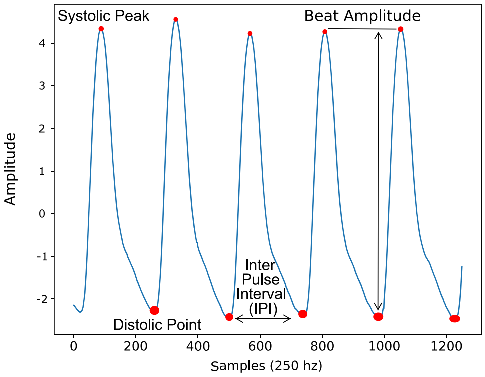

FIGURE 1. Points of interest in a PPG signal from the CapnoBase dataset include the systolic peak, which represents the peak of a heartbeat, and the diastolic point, which represents the lowest point in the heartbeat. The Inter-Pulse Interval (IPI) corresponds to the time between two consecutive diastolic points or systolic peaks, while the beat amplitude represents the difference between the systolic peak and the diastolic point.

modes for machine learning (ML) / Deep learning (DL) models.

A. PPG SIGNAL

The Photoplethysmogram (PPG) is a non-invasive optical method used to monitor changes in blood volume [14]. PPG offers a cost-effective way to measure various biological parameters, including heart rate, blood oxygen saturation, respiration rate [15], blood pressure [16], and blood glucose levels [17].

Figure 1 illustrates key points in a PPG segment from the CapnoBase dataset, highlighting the signal's periodic nature. Notably, PPG signals contain systolic and diastolic points, representing peak and trough values. The time between consecutive diastolic or systolic points, known as the inter-pulse interval (IPI), can be used to measure heart rate variability (HRV) [18]. Additionally, the difference between the diastolic point and the systolic peak can quantify beat amplitude. Due to its versatility, PPG signals have been widely employed in biometric authentication systems, a topic explored in detail in Section III.

B. USER AUTHENTICATION

User authentication models based on ML/DL are usually developed as one-class classification, binary classification, or multi-class classification. In one-class classification, the model is trained only on legitimate user data. This model is close to real-life scenarios, where adversary data are not available. Exemplary algorithms for this approach are local outlier factor, one-class support vector machine, and isolation forests. In binary classification, a model is trained with legit users' data and adversary data in an authentication problem, where the model classifies data as either honest or adversary. Both one-class classification and binary classification can be seen as authentication problems because the user claims the identity, and the model verifies the correctness of the user's claim. In multi-class classification mode, the model is trained with honest and adversary data in an identification problem, where the model identifies and authenticates a specific user out of an extensive set of users.

It is worth noticing that using binary or multi-class introduces a limitation. Specifically, the model needs to be trained on the legitimate user and adversary data, which is not usually available to the end-user. Also, using multi-class classification complicates maintaining the model, as the model would need to be updated and retrained with the addition or removal of users from the authentication system. On the contrary, one class classification provides the highest usability, as the model requires only the honest user data for the training process, and the model is user-specific. Hence, no expensive retraining operation is required for adding or removing users to/from the authentication system. However, one-class models usually provide lower performance than the binary or multi-class model, as the model is only limited to honest user data during the training phase, which makes the model prone to generalization problems like over-fitting or under-fitting.

In summary, since this work aims to provide a usable PPG authentication model, we limited our work to one-class classification algorithms, as they require only honest user data in training and need not to be retrained or modified after adding or removing users from the authentication system.

III. RELATED WORK

In this section, we explore the related work in the field of PPG-based authentication, categorizing methods into one-time and continuous authentication approaches.

A. PPG ONE-TIME AUTHENTICATION

PPG feature extraction methods can be categorized as fiducial or non-fiducial. Fiducial methods identify specific points like systolic peaks and diastolic points, while non-fiducial methods, such as wavelet transformations, and autoencoders, extract features without specific points.

1) FIDUCIAL PPG AUTHENTICATION

In the realm of one-time authentication, researchers have explored fiducial-based methods for PPG authentication. For instance, Sarkar et al. [27] introduced an authentication approach relying on converting fiducial points on the PPG signal to Gaussian representation and using Quadratic Discriminant Analysis to authenticate the user. Their model achieved an accuracy of 96% on the DEAP dataset. Shang and Wu [22] proposed a PPG-based authentication system that uses motion gestures extracted from PPG signals and uses Local Outlier Factor (LOF) to authenticate the user. Their model achieved an accuracy of 96.31% on their in-house dataset. Lovisotto et al. [21] developed a PPG authentication system utilizing mobile camera lenses, achieving an 8% Equal Error Rate (EER).

2) NON-FIDUCIAL PPG AUTHENTICATION

Conversely, non-fiducial methods have also garnered attention in PPG authentication. Choudhary and Manikandan [28]

TABLE 1. Summary PPG based authentication related works.

Reference	Authentication Mode	Dataset	users	Algorithm	Performance			
					EER	Accuracy	AUC	F1
Alotaiby et al.[19]	One Time	CapnoBase	42	Binary Random Forest	–	99.48 %	–	–
Hwang et al.[20]	One Time	BioSec2	100	Binary CNN-LSTM	2%	98%	–	–
Lovisotto et al.[21]	One Time	In-house	15	One-Class SVM	8%	–	–	–
Shang et al.[22]	One Time	In-house	12	One-Class LOF	2.3%	96.31%	–	–
Sancho et al.[23]	One Time	CapnoBase	42	One-Class Distance	1%	–	–	–
Yadav et al.[24]	One Time	DEAB	23	Binary SVM	2.11%	–	–	–
Luque et al.[25]	One Time	Torika	20	Binary CNN	–	–	83%	–
Karimain et al.[26]	One Time	CapnoBase	42	One-Class KNN	1.31%	99.84%	–	–
Sarkar et al.[27]	One Time	DEAP	23	Multi Class QDA	–	96%	–	–
Choudhary et al.[28]	One Time	MIT-BIH	30	One-Class NCC	0.29%	–	–	–
Pu et al[29]	Continuous	CapnoBase + BioSec1. Lab + In-house	120	Distance Measurement	5.5%	98%	–	–
Zhao et al.[4]	Continuous	CapnoBase	20	Binary Gradient Boosting	–	96%	–	–
Wu et al.[30]	Continuous	In-house	40	One-Class SVM	–	98.5%	–	86.67%

proposed a noise-robust PPG authentication scheme based on pulsatile waveform correlation, achieving an EER of 0.29%. Karimian et al. [26] harnessed non-fiducial features from PPG signals, outperforming fiducial point features with an accuracy of 99.84% and an EER of 1.31%. Luque et al. [25] devised a CNN-based authentication model, achieving an impressive Area Under the Curve (AUC) score of 83.2%. Sancho et al. [23] experimented with various non-fiducial feature extraction techniques and achieved an EER of 1% for the CapnoBase dataset. Yadav et al. [24] employed Continuous Wavelet Transform for authentication, yielding a mean EER of 2.11% on the DEAB dataset. Hwang et al. [20] combined CNN and LSTM networks, achieving a 98% accuracy rate on the BioSec2 dataset. Lastly, Alotaiby et al. [19] proposed an authentication model based on both PPG and ECG signals, achieving an impressive average accuracy of 99.48% on the CapnoBase dataset.

B. PPG CONTINUOUS AUTHENTICATION

In the realm of continuous authentication, several noteworthy approaches have been explored. Wu et al. [30] introduced a sensor-based continuous authentication system that leveraged SVM, autoencoders, and KNN models. Their model achieved an F1 score of 81.67% in the walking state. Zhao et al. [4] developed a continuous authentication system using custom PPG sensors, achieving a 90% accuracy rate. They also tested the system's resistance to PPG signal drifting over time, which indicated the need for periodic model updates to maintain accuracy. Pu et al. [29] proposed a continuous authentication scheme based on PPG signals, achieving a 5.5% EER for authentication.

In summary, existing PPG-based authentication methods have demonstrated promising results in both one-time and continuous authentication scenarios. However, they often encounter challenges related to PPG signal drifting over time and the duration of the registration phase. This paper introduces a novel session-based PPG authentication approach, which mitigates these challenges by proposing a session-based approach, reducing the impact of PPG drifting, and allowing for shorter registration phases. Additionally, dynamic model updating is employed to adapt to in-session PPG drifting.

IV. METHODOLOGY

In this section, we will detail the technical aspects of our proposed model. To start, we will provide an overview of the model. Next, we will examine the details of the template creation process, introducing both the proposed autoencoder model and the baseline model. After that, we will thoroughly discuss the authentication model, including the various one-class classification algorithms employed and the continuous updating scheme. We will also introduce the benchmarking datasets and explain our data preprocessing procedures. Our validation methodology will be presented, and we will conclude by introducing the evaluation metrics and describing the experimental testbed.

A. SYSTEM OVERVIEW

Figures 2 and 3 offer an overview of the registration and authentication phases, respectively, for our proposed system. In the registration phase (Figure 2), after passing the initial authentication system, the user's PPG signals are processed by an AutoEncoder (AE) to create a baseline template. A Local Outlier Factor (LOF) model is then trained on this template for future user authentication.

The authentication phase (Figure 3) involves continuous PPG data collection during user interactions. Simultaneously, an AE derives the authentication template, compared to the baseline using the LOF model. If both templates are sufficiently similar, the session continues with the LOF model updating. If not, the system locks. This two-step process occurs at each user session start, ensuring a secure authentication experience. Algorithm 1 shows an overview of the proposed authentication scheme.

The session-based nature of our proposed system was designed specifically to overcome the inter-session PPG drifting. Additionally, intra-session PPG drifting is managed through dynamic baseline template updates based on live PPG data. Also, the utilization of AE for unsupervised feature extraction, making it usable with only honest user data,

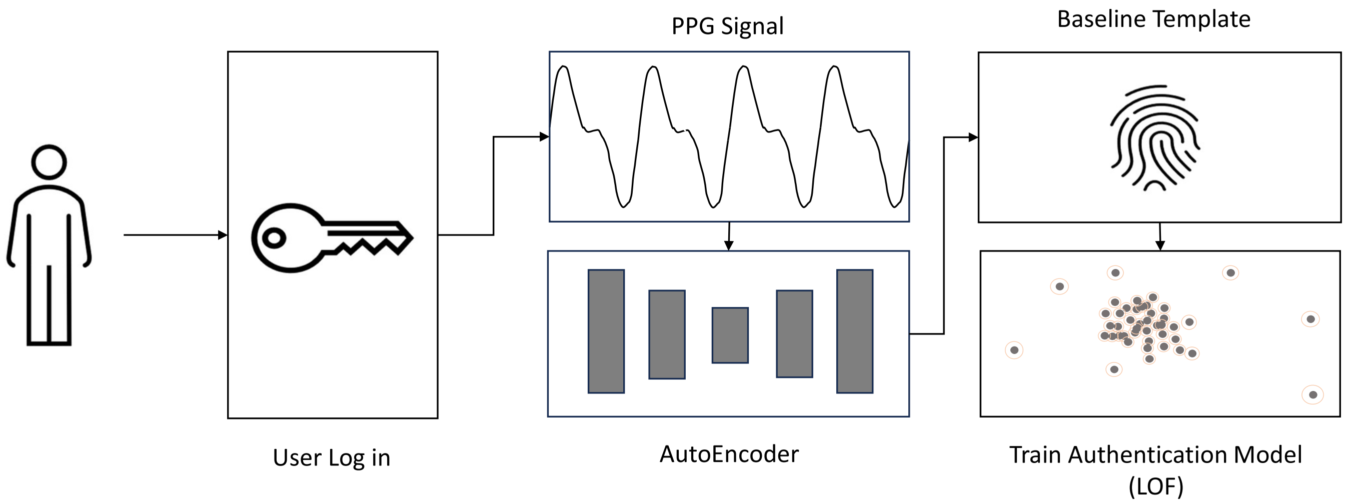

FIGURE 2. This diagram provides an overview of the Registration Phase within the proposed session-based authentication scheme. After successfully passing through the initial authentication layer (e.g., password-based system), PPG signals are collected and subsequently input into the autoencoder (see Section IV-B) to generate the baseline template. This baseline template is then employed to train the LOF authentication model (see Section IV-C1).

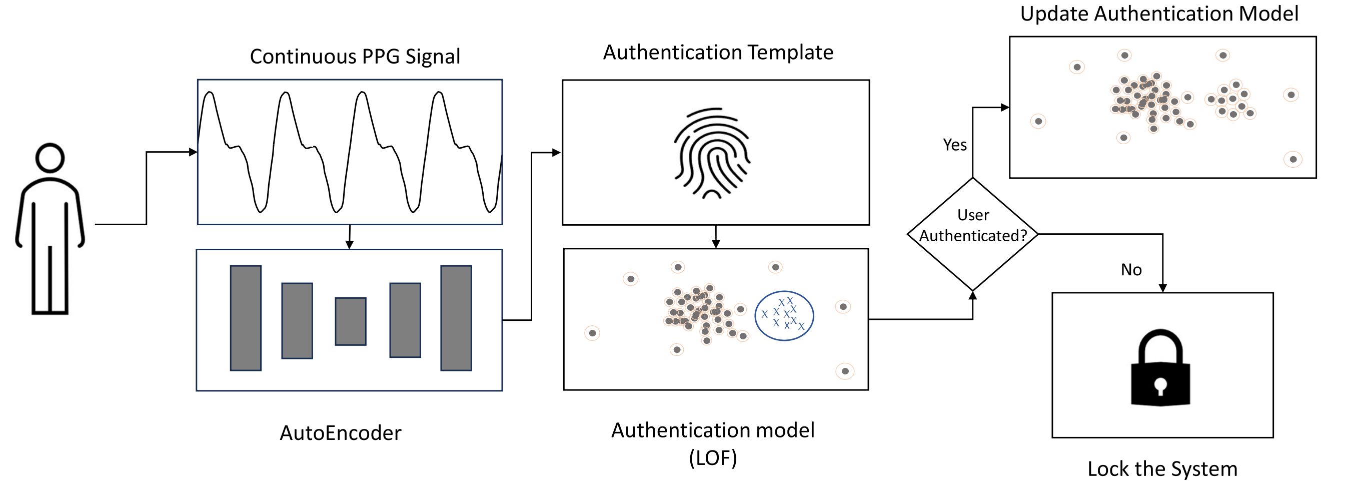

FIGURE 3. This diagram provides an overview of the Authentication Phase in the proposed session-based authentication scheme. Throughout the user session, continuous collection of PPG data occurs. Subsequently, the autoencoder (see Section IV-B) is employed to generate the authentication template. This template is utilized by the authentication model (see Section IV-C1) for user authentication. If the user is successfully authenticated, the authentication model is updated (refer to Section IV-C2). In case of unsuccessful authentication, the system is locked.

eliminating the need for an adversary dataset. Moreover, it's a one-size-fits-all solution, requiring just one AE for all system users (including new system users without the need to retrain the AE), reducing registration and update time and data demands.

In the following sections, we will discuss the details of the proposed authentication scheme.

B. TEMPLATE CREATION

Template creation is essentially a feature extraction challenge, aiming to derive a set of features that best represent the user. As demonstrated earlier, this can be achieved through fiducial or non-fiducial approaches. However, for our specific task, non-fiducial methods have been shown to outperform fiducial ones [26], so our focus in this work is on non-fiducial methods.

In this section, we introduce our proposed feature extraction method using Deep Autoencoders and also present our baseline method based on Wavelet Transformation (DWT) and Kernel Principal Component Analysis (KPCA).

1) AUTOENCODER

Autoencoders are widely used in unsupervised learning, feature extraction, and dimensionality reduction [29]. They aim to reproduce their input at the output layer while passing the data through a lower-dimensional bottleneck layer that encourages the model to keep relevant information and discard irrelevant details. The autoencoder consists of two

Algorithm 1 Authentication Algorithm

```
Data: PPG
Result: Authentication Decision
baseline_tmpt ← None;
while True do
    beats ← preprocess(PPG) ; /* filtering
        and segmentation */
    auth_tmpt ← encoder(beats);
    if user_tmpt is None then
        baseline_tmpt ← auth_tmpt;
    else
        decision ←
          detect_outliers(auth_tmpt, baseline_tmpt);

        if decision is normal then
            append(baseline_tmpt, auth_tmpt);
        else
            exit_session() ;
            break;
        end
    end
end
```

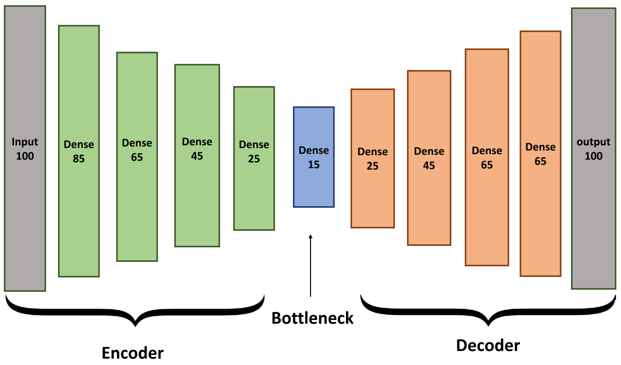

FIGURE 4. The design of the proposed Autoencoder: The autoencoder consists of four encoding layers, a bottleneck layer, and four decoding layers.

parts: an encoder E that maps input x to a code c, and a decoder D that reconstructs the output r from the code c. The error of the autoencoder is measured by calculating the distance between the input x and the output r [32]

Setting the dimensionality of the code c lower than that of x forces the autoencoder to find a compressed representation of x that minimizes the error, focusing on informative features for more accurate reconstruction. Autoencoders can be either shallow, with a single layer for the encoder and decoder, providing low computation cost and reasonable performance, or deep, with a stack of deep neural network layers, enabling more efficient learning of nonlinear patterns in the data, albeit with higher computational cost [32].

Our proposed deep autoencoder is composed of four fully connected layers representing the encoder part, one bottleneck layer, and four fully connected layers representing the decoder part. All the layers use a leaky ReLU activation function [33], with a batch normalization used for each layer to normalize the output of the layer, helping to create a more stable model [34]. The number of nodes in the encoding layers follows a decreasing order from 100% of the input shape at the input layer, then 85%, 65%, 45%, 25% for the first, second, third, and fourth layer of the encoder, respectively, then a bottleneck layer with 15% of the input shape, and then a decoder layers with the inverse number of nodes of the encoder layer, that is 25%, 45%, 65%, and 85%. Finally, an output layer of the same size as the input layer. The model is trained using mean square error loss function, and ADAM optimizer. Figure 4 shows the used autoencoder design.

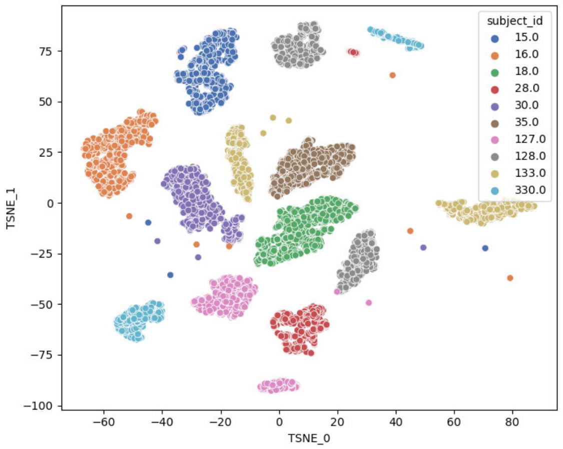

FIGURE 5. t-SNE [31] representation of the output of the encoder for PPG data that were not used for training—data from the CapnoBase dataset. The axes in the figure represent the two components from t-SNE.

The autoencoder is then trained over 500 epochs with a learning rate of 0.001 and a batch size of 500 beats. The model uses early stopping if the validation loss does not improve for 25 epochs. After the autoencoder is trained, the decoder part is discarded, and the encoder part is saved. Then, the encoder part is used to construct the user signature (named baseline template) through processing of the raw PPG signals.

Figure 5 shows a 2D projection of the encoder output using t-SNE [31] for the users' data that were not used to train the autoencoder—on CapnoBase dataset. The figure shows that the encoder can efficiently identify unique features from unseed raw PPG signal that can help to discriminate users.

2) DWT + KPCA

To assess the performance of our proposed system, we implemented an authentication model based on the approach introduced by Karimian et al. [26]. This model leverages Discrete Wavelet Transformation (DWT) and Kernel Principal Component Analysis (KPCA) for template creation. It's worth noting that their model stands out as one of the state-of-the-art solutions in PPG biometric authentication on the CapnoBase dataset, as highlighted in recent studies [35], [36], [37]. Their best-performing model achieved remarkable results with an average accuracy of 99.84% and an EER of 1.31% on the CapnoBase dataset [38].

Discrete Wavelet Transformation (DWT) is a signal processing technique that decomposes a signal into different frequency components, known as wavelet coefficients. It offers a multi-resolution analysis, capturing both high and low-frequency components of the PPG signal. DWT can effectively extract relevant features that represent the variations and patterns in the PPG signal. However, the coefficients of the DWT are sometimes longer than the signal itself. Thus, there is a need for a method that can reduce the size of this signal.

Kernel Principal Component Analysis (Kernel PCA) is a nonlinear extension of the traditional PCA method. It leverages a kernel function to map the data into a higher-dimensional feature space, where it becomes linearly separable. Kernel PCA and related techniques [39] can uncover complex patterns and relationships within the PPG data, enabling discrimination between different users and improving the performance of the authentication model.

For the implementation of the DWT+KPCA system, we used the same parameters as proposed by the original authors [26]. Specifically, we utilized the coif wavelet for DWT and the RBF kernel with 10 components for KPCA. This ensured a fair comparison between the two feature extraction methods in our study.

In their original work [26], Kaimian et al. employed a k-s test-based correlation filter for feature selection. However, due to a lack of details regarding their exact methodology, we opted for a mutual information-based feature selection approach in our implementation. This involved selecting the top 150 features that carried the most relevant information. To maintain a fair comparison between the DWTKCAP and AE models, we standardized the classification method, choosing the Local Outlier Factor algorithm, which demonstrated superior performance, as indicated in the results section.

To validate the accuracy of our implementation, we followed the original work's testing protocol, conducting 50 repetitions and averaging the outcomes. Our replicated model achieved impressive results, with an EER of 1.46% and an accuracy of 98.95%, closely aligning with the original work's reported metrics (99.84% accuracy and 1.31% EER).

C. AUTHENTICATION

After extracting features from the PPG signals using the template creation method, our next step involves user authentication through the authentication model. To tackle this challenge, we have chosen to employ the one-class classification model. The rationale behind this choice is that these models rely solely on honest user data for training, which enhances their practicality and ease of use. Additionally, we have implemented a continuous updating mechanism to address the issue of intra-session PPG signal drift. In the following subsections, we will delve into the specifics of both the authentication model and the continuous updating mechanism.

1) AUTHENTICATION MODEL

Our proposed solution relies on one-class classifier algorithms, which offer significant practicality, as they exclusively demand data from legitimate users for the training phase. To assess the effectiveness of our approach, we conducted experiments with various one-class classification algorithms, namely Local Outlier Factor (LOF) [40], One-Class Support Vector Machine (OCSVM) [41], Elliptic Envelope (EE) [42], and Isolation Forest (IF) [43]. We implemented all these algorithms using the scikit-learn Python library [44]. In each case, we adhered to the default parameters recommended by the scikit-learn library, except for LOF, where we adjusted the number of neighbors to 30 beats, aligning it with half of the buffer size for optimum performance.

During the registration phase, the one-class classification algorithm is first trained on the baseline template. Then, during the authentication phase, it continuously tests the similarity between the authentication and baseline templates. Since both templates are composed of non-fiducial features representation of the single heartbeats, this problem can be rephrased as outlier detection between the authentication template and the baseline template. If the majority of beats in the authentication template are considered outliers to beats in the baseline template, it indicates the existence of an adversary. However, if most beats are considered inliers, we conclude that the subject under test is the legitimate user.

The performance of our authentication model is influenced by two critical parameters: the size of the baseline template, which is controlled by the buffer size, and the configuration of the authentication template, which is regulated by the window size. The buffer size, which determines the size baseline template, governs the amount of data available for training the model. For instance, with a buffer size of 60 beats, the model is initially trained on data of 60 user beats. While increasing the buffer size tends to improve the model's quality, it does come at the cost of a longer active training period.

Conversely, the length of the authentication template can be tailored by adjusting the window size. For example, setting a window size of 10 beats means the system attempts to authenticate a user every 10 beats. This configuration facilitates early detection of potential adversaries. However, increasing the window size (e.g., to 60 beats) can enhance accuracy but slows down the detection process. In the results section, we conduct analysis on the effect of varying both the buffer and window sizes.

2) CONTINUOUS UPDATE

The continuous update of the one-class classification algorithm is performed by adding the authentication template to the baseline template if it is detected as inlier, then retraining the one-class classification algorithm again on the updated baseline template—the rate of this process can be tuned to fit the system requirement. In detail, the continuous updating process can be performed after every M authentication cycle, where M is chosen by the system admin according to the environment where the model would be running. For example, on a mobile device with limited power usage, M can be large to use the least amount of power. However, on a desktop, M can be as small as 1, which means the baseline template would be updated after every authentication operation. In the testing phase in this paper, M is set to 1.

This continuous update of the authentication model makes the model more robust against intra-session PPG drifting—drifting being tied to the user's mental or physical state. An interesting feature of this model is also its self-healing property: the model would recover its usefulness even if it was partially corrupted by an adversary, as long as the honest user has more physical presence over the adversary.

D. BENCHMARKING DATASETS

In our evaluation, we utilized three publicly available datasets: Bidmc [45], CapnoBase [38], and BioSec2 [20] datasets. The Bidmc dataset comprises 8 minutes of PPG data extracted from the MIMIC II matched waveform database. This data was collected from 53 patients in the intensive care unit during a single session. The CapnoBase dataset contains 8 minutes of PPG recordings from 43 users, including 29 pediatric and 13 adult subjects, all recorded during resting conditions in a single session. Furthermore, to measure the impact of our proposed model on the inter-session PPG drifting, we used the BioSec2 dataset, which consists of two sessions (17 days apart) of PPG recordings from 100 subjects. Each session comprises three trials, each lasting for 90 seconds. These trials are designed to model the randomness of data collection from the sensing device. In the next subsection, we are going to introduce the preprocessing stage that was applied to all our used datasets.

1) PREPROCESSING

Our proposed scheme commences with the initial step of cleansing the received PPG signal to eliminate static noise interference. To achieve this, we employed the PPG_clean function from the neurokit2 library [46], which builds upon the framework introduced by Elgendi [7]. This function employs a third-degree Butterworth bandpass filter with frequency cutoffs set at 0.5 Hz and 8 Hz. This filtering process effectively eliminates noise stemming from motion artifacts and baseline drift [7]. Subsequently, we segment the PPG signal into distinct beats utilizing a state-of-the-art beat segmentation algorithm developed by Elgendi et al. [47]. This algorithm is implemented in the neurokit2 Python library [46].

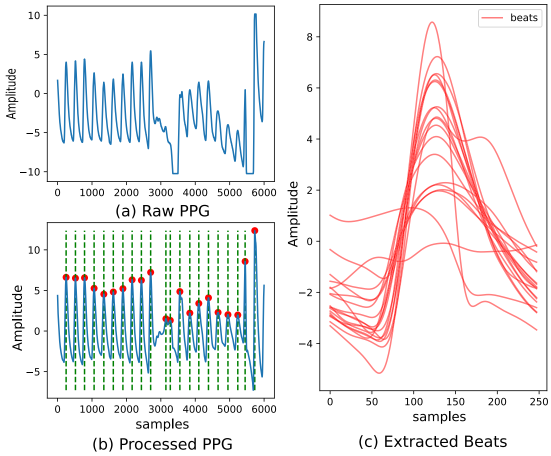
(a) Raw PPG
(b) Processed PPG
(c) Extracted Beats

FIGURE 6. Illustration of a noisy segment from the CapnoBase dataset. The graph shows the three stages of preprocessing: (a) raw signal; (b) filtered signal and marking of the places of the systolic peaks; and, (c) extracted beats from the PPG signal.

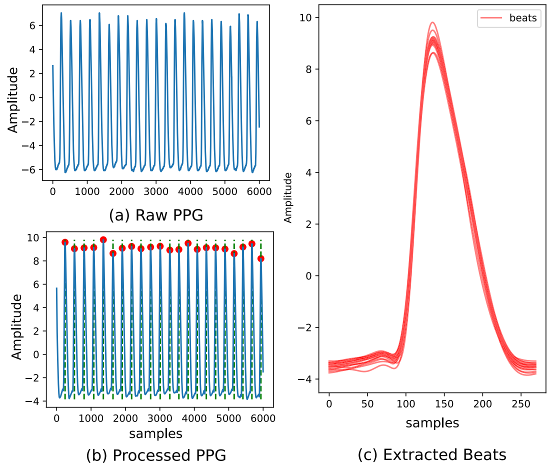
(a) Raw PPG
(b) Processed PPG
(c) Extracted Beats

FIGURE 7. Illustration of a clean segment from the CapnoBase dataset. The graph shows the three stages of preprocessing: (a) raw signal; (b) filtered signal and marking of the places of the systolic peaks; and, (c) extracted beats from the PPG signal.

Figures 6 and 7 showcase segments from the CapnoBase dataset, illustrating both the noisy and clean PPG signals. It is evident from these figures that the peak detection algorithm successfully identifies the systolic peaks within the PPG signal, both in the presence of noise and in the clean signals. However, it is worth noting that the filtering technique does not completely eliminate all signal noise, underscoring the importance of devising a robust algorithm capable of operating effectively under noisy conditions.

E. VALIDATION METHODOLOGY

The validation process serves a twofold purpose: first, to assess the template creation models' ability to generalize to unseen data, and second, to evaluate the overall performance of the authentication system. To achieve this, we partition all datasets into two distinct sets: the encoder dataset and the authentication dataset. The encoder dataset is employed to train the template creation model, whether it is an autoencoder or a baseline model, while the authentication dataset is reserved for assessing authentication performance.

To evaluate the generalizability of the template creation models, we adopted an unsupervised training approach. Initially, these models are trained on the encoder dataset, and subsequently, the data used for the encoder dataset is discarded and not used in the subsequent evaluation step. This process simulates a scenario where a single template creation model is trained for an organization on a subset of users and then applied to all users in the organization. This approach enhances usability, as it eliminates the need for retraining when adding or removing users from the authentication system.

After training the template creation models, we proceed to evaluate authentication performance. In this evaluation, we employ a one-vs-all cross-validation approach on the authentication dataset. This involves selecting one user as the honest user while considering all other users as potential adversaries. However, this approach can lead to a significant class imbalance problem. To address this, we randomly sample a subset of data from the adversary users, ensuring it matches the length of the honest user's data, effectively balancing the dataset.

It is essential to note that not all of the honest user's data is utilized during training. Only the initial n beats, where n represents the buffer size (baseline template), are used for training, while the remaining data is reserved for testing. For instance, if the buffer size is set to 60, the first 60 beats are used for training the authentication model, and the remaining beats are used for testing. Additionally, in the case of the BioSec2 dataset, which contains multiple trials for each session (unlike the other two datasets), the evaluation process is repeated for each user and trial. This approach aligns with our session-based model, where each trial represents the start of a different session.

Finally, we repeat the entire simulation 50 times for each dataset, with users randomly selected for training the template creation models and the authentication model. Our experiments have shown that a buffer size of 60 beats and a window size of 10 beats provide the best overall performance for both the CapnoBase and BIDMC datasets. As for the BioSec2 dataset, due to its limited data duration, we used a buffer size of 30 beats. These choices were made based on their ability to deliver optimal performance while maintaining a balance between accuracy and waiting time, as we will discuss in detail in the results section.

F. EVALUATION METRIC

Evaluating authentication models and classification models in general depends on the number of True positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) samples. To describe the system performance based on those characteristics, the following evaluation metrics have been used.

- **False Acceptance Rate**: This is the rate at which an honest user is denied access to the system. FRR can be calculated as following:

$$\frac{\mathsf{FP}}{\mathsf{FP}+\mathsf{FN}} \tag{1}$$

- **False Rejection Rate**: The rate at which an adversary is given access to the system

$$\frac{\mathsf{FN}}{\mathsf{TP}+\mathsf{FN}} \tag{2}$$

- **Equal Error Rate**: is a method to optimize the trade-off between FRR and FAR—both of them should be as low as possible. EER is calculated by plotting FAR and FRR on a ROC plot, where the lowest point in the curve represents EER.
- **Precision**: is the probability that a user is being given access to the system is honest.

$$\frac{\mathsf{TP}}{\mathsf{TP}+\mathsf{FP}} \tag{3}$$

- **Recall**: is the probability that an honest user would be given access to the system.

$$\frac{\mathsf{TP}}{\mathsf{TP}+\mathsf{FN}} \tag{4}$$

- **F1**: F1 is defined by the harmonic mean of the Precision and Recall of the model, which emphasizes a balance between Precision and Recall.

$$2 \times \frac{\mathsf{Precision} \times \mathsf{recall}}{\mathsf{precision} + \mathsf{recall}} \tag{5}$$

- **Authentication Accuracy**: is the number of correctly identified instances over the total number of instances.

$$\frac{\mathsf{TP}+\mathsf{TN}}{\mathsf{TP}+\mathsf{FP}+\mathsf{TN}+\mathsf{FN}} \tag{6}$$

G. TESTBED

The experiment was done on ubuntu 18.04.4 with a Xeon Gold 6128 CPU, Quadro P4000 8GB GPU with CUDA version 10.1, and using Python-3.9.7. For developing the autoencoder, we used TensorFlow 2.4.1. And, for all the outlier detection algorithms, we used sci-kit learn 0.24.2. As for performing beats extraction, we used biopsy 0.7.3. Furthermore, we used Scipy 1.6.2 and neurokit2 0.1.4.1 to filter the signal, and DWT implementation in pywavelete 1.1.1 was used in the baseline model.

V. RESULTS AND DISCUSSION

In this section, we will comprehensively assess the performance of our proposed authentication model and authentication scheme. To do so, we will begin by examining single-session scenarios and subsequently extend our analysis to multiple-session settings. Additionally, we will delve into the interpretability of the model's decisions using GRAD-CAM models. We will then investigate the impact of various model parameters, including the choice of authentication algorithms, buffer size, window size, and decision thresholds. Furthermore, we will evaluate the effectiveness of the automatic recovery feature embedded in our model. Finally, we will conclude this section by comparing our findings with related work in the field.

A. SINGLE SESSION SETTINGS

Table 2 shows the performance of the baseline model (DWT+KPCA) and our proposed model (AE) Under single session settings on both the BIDMC and CapnoBase datasets. The results illustrate that the proposed model can provide lower EER and higher accuracy and F1 over the baseline model. For the BIMDC, the proposed model provided a reduction in EER by 2.3%, and in the CapnoBase, the proposed model provided an increase of more than 6% in F1 score. Furthermore, both models achieved almost perfect precision on the CapnoBase dataset, whereas the proposed model achieved a higher recall with an increase of more than 6%. Also, on the BIMDC dataset, the proposed model achieved better performance than the baseline model across all metrics.

Additionally, figures 8 and 9 depict the comparison of the kernel density estimate (KDE) distribution for EER, F1, and accuracy between the proposed and baseline models on the CapnoBase and BIDMC datasets, respectively. The figures indicate that the proposed model exhibits a higher density around small EER values in comparison to the baseline model. Also, a similar trend is observed for F1 and accuracy in both datasets. Specifically, the proposed model has a higher density around high score areas compared to the baseline model. This finding suggests that the proposed model is more likely to perform better than the baseline model based on the evaluated datasets.

B. MULTIPLE SESSION SETTING

The results presented in Table 3 offer a comparison of our session-based authentication scheme, examined across multiple sessions spanning 17 days, in reference to Hwang et al.'s study [20]. In Hwang et al.'s scheme, the objective was to establish a stable representation of the PPG signal that could withstand PPG drifting. When their model was assessed across multiple sessions, it yielded an Equal Error Rate (EER) of 12.9% and an accuracy of 87.1%. Notably, in a single session, their system achieved an EER of 2% and an accuracy of 98%. This resulted in an 11% reduction in accuracy and a 10.9% increase in EER, highlighting the detrimental impact of inter-session PPG drifting on their authentication performance.

In contrast, our proposed scheme demonstrates a substantial performance improvement. We achieved an EER of 2.7% and an accuracy rate of 95.9%, showcasing our system's robust ability to accurately authenticate users across multiple sessions. This demonstrates the effectiveness of our proposed scheme in mitigating the adverse effects of inter-session PPG drifting while maintaining performance levels comparable to the single-session scenario.

C. MODEL EXPLAINABILITY

Furthermore, to better understand how the Autoencoders can extract effect representation of PPG we employed Gradient-weighted Class Activation Mapping (GRAD-CAM) algorithm [48]. This algorithm can identify the important parts in PPG beats that have the most significant effect on the Autoencdoer output. GRAD-CAM calculates the average weights of the Dense layers based on the inferred class probabilities, allowing us to pinpoint the segments of the input that had the greatest impact on the model's decision. Figures 10 and 11 illustrate the areas of the signal that contribute the top 25% to the Autoencoder output for both the CapnoBase and BIDMC datasets. The figures show an average and standard deviation of 20,000 randomly selected beats from each dataset. It is evident that the AE model tends to emphasize the area around the systolic peak of the signal, and the diastolic points at the start and end of each signal also provide significant information to the AE model. This suggests that the proposed AE model automatically learns to focus on important areas of interest and can ignore parts of the signal that do not provide much information.

D. EFFECT OF SYSTEM PARAMETERS

1) OUTLIER DETECTION ALGORITHM

Additionally, we conducted an evaluation to assess the impact of various outlier detection algorithms, which included the following: Local Outlier Factor (LOF) [40]; One-Class Support Victor Machine (OCSVM) [41]; Elliptic Envelope (EE) [42]; and Isolation Forest (IF) [43]. We used the BIMDC dataset in this test, where we fixed the random splitting parameters and only changed the outlier detection algorithm with every run, using a window size of 10 beats. Table 4 shows that the best algorithm in the majority of the metrics is the LOF, except for the FAR, where SVM has the lowest FAR. Also, the fastest algorithm to run the full experiment was the SVM, followed by LOF with a difference of around 6 seconds.

2) EFFECT OF DECISION THRESHOLD

In the majority voting process, a crucial parameter to consider is the decision threshold. This threshold plays a key role in determining whether the authentication template is considered valid. Specifically, if the percentage of normal beats in the authentication template surpasses a certain threshold, the template is authenticated as belonging to a normal user. The authentication threshold can be configured by the system administrator to prioritize either stringent security (by setting a high authentication threshold), which reduces the False Acceptance Rate (FAR) at the expense of increasing the False Rejection Rate (FRR), or to provide a better user experience (by choosing a lower authentication threshold) with reduced security (resulting in a lower FAR).

Table 5 demonstrates the impact of varying the threshold value from 0.1 to 0.9 using the LOF model on the BICMD dataset. The experiment reveals that a threshold value of

TABLE 2. Authentication performance results.

Data set	Model	EER	FRR	FAR	Accuracy	F1	Precision	Recall
CapnoBase[38]	Baseline (Karimian et al. [26])	2.1 ±4.8	18.8 ±28.3	0.5 ±1.7	90.3 ±13.5	85.0 ±24.3	98.8 ±5.1	80.1 ±28.4
	Proposed	**2.0** ±4.8	**13.2** ±19.0	**0.4** ±1.9	**93.5** ±8.9	**91.3** ±13.7	**99.5** ±2.0	**86.7** ±19.0
BIDMC [45]	Baseline (Karimian et al. [26])	8.1 ±12.7	11.8 ±14.3	8.0 ±18.2	90.2 ±12.0	89.6 ±12.1	93.6 ±13.4	88.1 ±14.4
	Proposed	**5.8** ±11.3	**11.7** ±14.5	**4.9** ±14.8	**91.8** ±11.2	**91.0** ±12.0	**95.9** ±11.3	**88.3** ±14.5

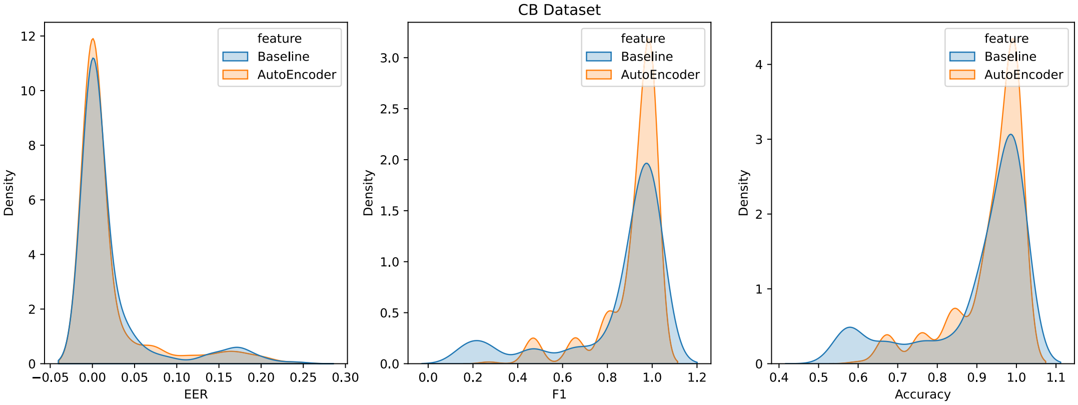

FIGURE 8. Kernel density estimate (KDE) plot comparing the distribution of Equal Error Rate (EER), F1 score, and Accuracy for the proposed and baseline models on the Capnobase dataset.

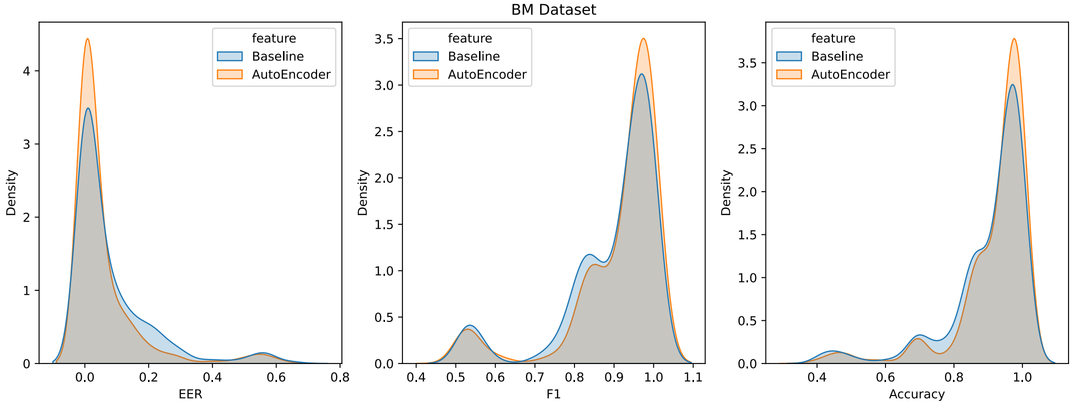

FIGURE 9. Kernel density estimate (KDE) plot comparing the distribution of Equal Error Rate (EER), F1 score, and Accuracy for the proposed and baseline models on the BIDMC dataset.

TABLE 3. Evaluation of the proposed authentication scheme under multiple session settings with time lapses of 17 days using the BioSec2 dataset.

Scheme	EER	Accuray	F1	FRR	FAR
hwang et al. [20]	12.9	87.1	-	-	-
Ours	2.7	95.9	95.9	1.2	6.3

TABLE 4. The effect of different outlier detection algorithms.

Algorithm	Accuracy	F1	FAR	FRR	Duration (seconds)
Elliptic Envelope	88.14	86.41	9.83	14.04	575.49
Isolation Forest	87.89	85.96	8.23	16.34	372.20
Local Outlier Factor	91.49	90.49	5.38	11.91	18.21
one class SVM	61.67	31.23	0.06	80.22	12.98

0.5 strikes a balance between FAR and FRR. As the threshold is increased, we can anticipate a general rise in FAR, at the cost of increasing FRR. Conversely, reducing the threshold prioritizes FRR over FAR.

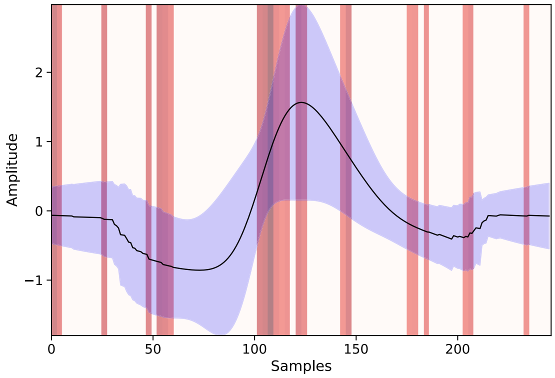

FIGURE 10. Grad-CAM analysis on the BIDMC dataset: Highlighting the Top 25% contributors to model decision.

3) WINDOW SIZE

For testing the effect of window size, we tested different window sizes of 1, 5, 10, 20, 30, 60 beats, respectively. Table 6

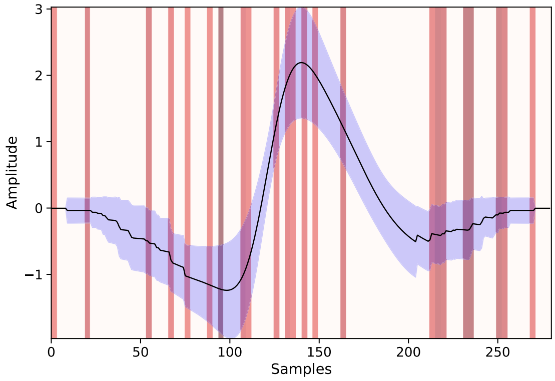

FIGURE 11. Grad-CAM analysis on the CapnoBase dataset: Highlighting the Top 25% contributors to model decision.

TABLE 5. Effect of different threshold value for the majority voting process.

Threshold	F1	ACC	FAR	FRR
0.1	90.84	88.97	19.91	1.28
0.2	90.89	89.75	16.51	3.37
0.3	92.88	92.61	9.78	4.75
0.4	93.71	93.95	5.35	6.80
0.5	92.20	92.71	6.07	8.62
0.6	91.68	92.63	3.71	11.38
0.7	88.34	90.06	4.16	16.27
0.8	84.94	88.30	1.58	22.76
0.9	77.58	83.12	2.66	32.45

TABLE 6. The effect of window size on BIMDC dataset.

window	Accuracy	F1	FAR	FRR
1	89.82	88.99	5.64	15.14
5	90.72	90.18	7.05	11.71
10	90.09	89.60	8.63	11.31
20	92.97	92.28	5.16	9.05
30	93.67	92.93	5.00	7.75
60	95.94	95.35	3.74	4.41

shows the performance of the model on different window sizes on the BIMDC dataset. According to the table, the model's performance improves as the window size increases. Moreover, the table illustrates that the increase in window size leads to a reduction in both the false acceptance rate (FAR) and false rejection rate (FRR). Moreover, determining the optimal window size is not solely based on the performance of the model, since larger window sizes may result in slower detection of adversaries than models with smaller window sizes. As a result, selecting the optimal window size is dependent on the system's requirements and characteristics.

4) BUFFERING DURATION

Our proposed model requires a buffering time at the start of each session, where the model collects enough data to create the baseline template. At this buffering stage, the continuous authentication system would be inactive. Table 7 shows the effect of changing the registration buffer size on the performance of the model using the CapnoBase dataset. From the table, we could see that the increase in buffer size resulted in small increase in accuracy and a small reduction in FAR. These results suggest that the proposed model can effectively authenticate users with as few as 15 beats. Thus, it is reasonable to adapt session-based continuous authentication over traditional continuous authentication, as long-term PPG drifting between sessions can lead to an over a 10% increase in EER [20].

TABLE 7. The effect of different buffering sizes on the CapnoBase dataset.

Buffer Size	Accuracy	F1	FAR	FRR
15	92.44	92.07	4.94	10.21
30	92.45	91.94	4.72	10.51
45	92.57	91.85	4.42	10.62
60	92.01	91.34	5.39	10.83
75	92.25	91.30	4.89	10.94
90	93.00	91.84	3.51	11.00

E. AUTOMATIC RECOVERY

We have also evaluated the automatic recovery features of our proposed system against potential breaches by adversaries. In particular, via continuous updating. In detail, we evaluated the efficacy of continuous updating against the introduced scenario by training the LOF model with different levels of corruption, using a combination of user and adversary data with ratios of adversary data ranging from 10% to 50%. We compared the performance of the proposed model with and without continuous updating for the BIDMC dataset, and the results of our evaluation are presented in Table 8. Our findings demonstrated that the introduction of continuous updating significantly improved the resilience of the proposed model. In fact, we observed a performance boost, with an increase in accuracy of more than 9% and a reduction of FRR by more than 17%. Therefore, our study suggests that continuous updating can effectively improve the resilience of biometric authentication models.

F. COMPARISON WITH RELATED WORK

Comparison between our proposed model and related works can be analyzed from two perspectives: security and utility. In terms of security properties, our proposed approach focuses on addressing both intra-session and inter-session PPG drifting. In the work of Zhao et al. [4], the authors proposed a model retraining solution for inter-session PPG drifting, involving periodic updates to the user signature every 3 hours, which can be computationally expensive and increase waiting times for authentication. Instead, our proposed session only requires training of a small LOF model at the start of each session. This is possible as our system utilizes a heavy Autoencoder model, trained only once on a small subset of user data by the system administrator, along with a light Local Outlier Factor (LOF) model trained privately on the user data. During each session, only the LOF model is trained, requiring as few as

TABLE 8. Overview of the effect of continuous updating on the BIDMC dataset with different corruption levels.

Corruption Rate	Continuous Updating	F1	Accuracy	Precision	Recall	FAR	FRR	EER
0.1	FALSE	84.74	85.05	87.86	87.25	12.75	16.96	10.14
	TRUE	89.48	89.81	93.12	88.23	11.77	8.69	9.73
0.2	FALSE	82.29	81.45	82.90	88.30	11.70	24.90	12.05
	TRUE	87.70	87.29	89.92	88.78	11.22	14.07	13.05
0.3	FALSE	81.85	80.78	82.59	88.09	11.91	26.03	12.69
	TRUE	85.67	84.78	86.40	88.33	11.67	18.54	15.10
0.4	FALSE	77.60	74.69	74.72	88.41	11.59	38.36	21.67
	TRUE	83.47	81.71	81.57	89.55	10.45	25.71	18.95
0.5	FALSE	79.68	77.63	78.89	88.17	11.83	32.38	16.99
	TRUE	87.41	86.97	89.06	89.00	11.00	14.91	13.56

TABLE 9. Comparison between different properties of the proposed authentication model and the related work.

Work	Resistance to PPG Drifting	Registration Duration	Authentication Duration	Data Needed for Training	Adversary Recovery
Wu et al. [30]	No	2.4 Hours	12 Seconds	Honest Only	No
Zhao et al. [4]	Model Retraining	300 Seconds	3 Seconds	Honest + Adversary	No
Pu et al. [29]	No	108 Seconds	36 Seconds	Honest + Adversary	No
Hwang et al. [20]	Time stable features	270 Seconds	90 Seconds	Honest + Adversary	No
karimian et al. [26] (Baseline)	No	240 Seconds	10 Beats	Honest Only	No
Ours	Session Based Authentication	15 Beats	20 Beats	Honest Only	Yes

15 beats, minimizing the response time for authentication. Furthermore, our model incorporates continuous updating for the lightweight LOF model with newly collected user data, providing a self-healing property that addresses intra-session PPG drifting and strengthens the model against potential adversarial breaches, as shown in Table 8. None of the compared related works targeting continuous authentication offer this self-healing capability, giving our proposed solution a significant advantage.

In terms of authentication duration, our model performs similarly to the compared related works, offering a rapid decision-making process. As the proposed model can provide an authentication decision with as low as a single beat, achieving an accuracy of around 90%. The optimal authentication duration, balancing performance and waiting time, is 20 beats with an accuracy of around 93%. This translates to a decision every 20 seconds, assuming a resting heart rate of 60 beats per minute for users. Also, our session-based authentication scheme eliminates the errors resulted from inter-session PPG drifting, which could reach in more than 10% increase in EER.

A striking feature of our proposed model is the low registration duration, which is the time required to collect data to generate a user's signature. Our solution can generate a user signature with as few as 15 beats, enabling quick initiation of each user session with minimal time overhead. In contrast, some compared works require up to 2.4 hours or a minimum of 108 seconds for registration.

One limitation of our solution is that the registration phase needs to be run at the beginning of each user session. Consequently, the continuous authentication system may not be activated until enough data is collected to generate the user signature. However, it is essential to note that the data required for the registration phase is relatively short, with a minimum of 15 beats. This duration is comparable to the authentication time in related works, where users often have to wait before receiving the initial results. As a result, the response time of our system remains comparable to that of related works, ensuring that the overall user experience is not significantly slower.

Additionally, many of the compared related works demand honest user data and adversary data to train their models, limiting the utility of the models. In contrast, our proposed solution requires data only for the honest user, making it more feasible to implement in real-world scenarios.

Thus, the proposed model can provide greater security while being more convenient than the related work, making it easier to implement and adapt it in real world scenarios. Comparing the performance of our proposed model directly with related works is challenging due to variations in datasets, evaluation folds, and specific evaluation schemes employed by different authors. To provide a sense of comparison, we implemented the solution proposed by karimian et al. [26] as our baseline, as discussed in Section IV-B2.

Our results, presented in Table 2, demonstrate a significant advantage of our proposed model over the karimian-based model. We achieved over a 6% increase in the F1 score on the Capnobase dataset, while also significantly reducing the execution time by 85%. These results highlight the superior performance and efficiency of our proposed model when compared to the selected baseline. Also, our model has been shown to eliminate the error increase due to PPG drifting which could be more than 10% increase in EER as shown in Table 3.

Furthermore, for execution speed, we empirically evaluated both solutions' speed. Generating an encoder representation of 18,000 beats takes around 0.5 seconds, whereas the baseline model representation for the same number of beats takes approximately 3.6 seconds with both solutions running on CPU mode.

VI. CONCLUSION AND FUTURE WORK

In this study, we present a novel session-based continuous authentication model utilizing PPG signals. Our approach leverages autoencoders (AE) to efficiently generate lower-dimensional representations of raw PPG signals by capturing their nonlinear relationships. Additionally, the model incorporates outlier detection algorithms, particularly the local outlier factor (LOF), for user authentication. The AE model is trained once on a subset of system users by the system admin and can then be used for all users without the need for retraining. As per the lightweight LOF model, it can efficiently initialize and authenticate a new user with a short buffer duration of as low as 15 beats. The session-based nature of our model ensures its immunity to inter-session PPG drifting, wherein user PPG signals may vary between different sessions.

The proposed model achieved an F1 score of 95.9%, 91.3% and 91.0% in the BioSec2, CapnoBase, and BIMDC datasets, respectively. Moreover, the continuous updating scheme helps the model recover from adversarial breaches and resist intra-session PPG drifting, resulting in an accuracy boost of over 9%. These results represent an improvement against state-of-the-art solutions. Our solution is also seven times faster at run time than competing state-of-the-art solution (while requiring more computation in terms of training—this latter overhead being sustained just once, and later amortized over the lifetime of the solution).

A future direction is to explore the effect of emotional states, physical states, and cardiovascular disease on the quality of PPG authentication. Also, the research community would benefit from a public PPG dataset collected from commercial devices with a long session duration (e.g., over 10 minutes). This would help in comparing the performance change between medical-grade and commercial devices for session-based PPG continuous authentication.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] P. A. Thomas and K. P. Mathew, "A broad review on non-intrusive active user authentication in biometrics," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 1, pp. 339–360, Jan. 2023.

[2] L. Hernández-Álvarez, J. M. de Fuentes, L. González-Manzano, and L. H. Encinas, "Privacy-preserving sensor-based continuous authentication and user profiling: A review," *Sensors*, vol. 21, no. 1, p. 92, Dec. 2020.

[3] S. Mondal and P. Bours, "A study on continuous authentication using a combination of keystroke and mouse biometrics," *Neurocomputing*, vol. 230, pp. 1–22, Mar. 2017.

[4] T. Zhao, Y. Wang, J. Liu, Y. Chen, J. Cheng, and J. Yu, "TrueHeart: Continuous authentication on wrist-worn wearables using PPG-based biometrics," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 30–39.

[5] W. Louis, M. Komeili, and D. Hatzinakos, "Continuous authentication using one-dimensional multi-resolution local binary patterns (1DMRLBP) in ECG biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2818–2832, Dec. 2016.

[6] F. Ahamed, F. Farid, B. Suleiman, Z. Jan, L. A. Wahsheh, and S. Shahrestani, "An intelligent multimodal biometric authentication model for personalised healthcare services," *Future Internet*, vol. 14, no. 8, p. 222, Jul. 2022.

[7] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiol. Rev.*, vol. 8, no. 1, pp. 14–25, Jun. 2012.

[8] L. Li, C. Chen, L. Pan, J. Zhang, and Y. Xiang, "SoK: An overview of PPG's application in authentication," 2022, *arXiv:2201.11291*.

[9] P. Spachos, J. Gao, and D. Hatzinakos, "Feasibility study of photoplethysmographic signals for biometric identification," in *Proc. 17th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2011, pp. 1–5.

[10] A. Bonissi, R. D. Labati, L. Perico, R. Sassi, F. Scotti, and L. Sparagino, "A preliminary study on continuous authentication methods for photoplethysmographic biometrics," in *Proc. IEEE Workshop Biometric Meas. Syst. Secur. Med. Appl.*, Sep. 2013, pp. 28–33.

[11] N. Karimian, Z. Guo, M. Tehranipoor, and D. Forte, "Human recognition from photoplethysmography (PPG) based on non-fiducial features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 4636–4640.

[12] S. Hinatsu, D. Suzuki, H. Ishizuka, S. Ikeda, and O. Oshiro, "Evaluation of PPG feature values toward biometric authentication against presentation attacks," *IEEE Access*, vol. 10, pp. 41352–41361, 2022.

[13] L. Li, C. Chen, L. Pan, J. Zhang, and Y. Xiang, "Video is all you need: Attacking PPG-based biometric authentication," in *Proc. 15th ACM Workshop Artif. Intell. Secur.*, Nov. 2022, pp. 57–66.

[14] R. Donida Labati, V. Piuri, F. Rundo, and F. Scotti, "Photoplethysmographic biometrics: A comprehensive survey," *Pattern Recognit. Lett.*, vol. 156, pp. 119–125, Apr. 2022.

[15] S. Singh, M. Kozlowski, I. García-López, Z. Jiang, and E. Rodriguez-Villegas, "Proof of concept of a novel neck-situated wearable PPG system for continuous physiological monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[16] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, and M. Fuchs, "Assessment of deep learning based blood pressure prediction from PPG and rPPG signals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 3815–3825.

[17] A. Y. Alhaddad, H. Aly, H. Gad, A. Al-Ali, K. K. Sadasivuni, J.-J. Cabibihan, and R. A. Malik, "Sense and learn: Recent advances in wearable sensing and machine learning for blood glucose monitoring and trend-detection," *Frontiers Bioeng. Biotechnol.*, vol. 10, May 2022, Art. no. 876672.

[18] G. Lu, F. Yang, J. A. Taylor, and J. F. Stein, "A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects," *J. Med. Eng. Technol.*, vol. 33, no. 8, pp. 634–641, Nov. 2009.

[19] T. N. Alotaiby, S. A. Alshebeili, G. Alotibi, and G. N. Alotaibi, "Recurrence quantification analysis for PPG/ECG-based subject authentication," in *Proc. 4th Int. Conf. Data Intell. Secur. (ICDIS)*, Aug. 2022, pp. 288–291.

[20] D. Y. Hwang, B. Taha, D. S. Lee, and D. Hatzinakos, "Evaluation of the time stability and uniqueness in PPG-based biometric system," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 116–130, 2021.

[21] G. Lovisotto, H. Turner, S. Eberz, and I. Martinovic, "Seeing red: PPG biometrics using smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 3565–3574.

[22] J. Shang and J. Wu, "A usable authentication system using wrist-worn photoplethysmography sensors on smartwatches," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2019, pp. 1–9.

[23] J. Sancho, Á. Alesanco, and J. García, "Biometric authentication using the PPG: A long-term feasibility study," *Sensors*, vol. 18, no. 5, p. 1525, May 2018.

[24] U. Yadav, S. N. Abbas, and D. Hatzinakos, "Evaluation of PPG biometrics for authentication in different states," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 277–282.

[25] J. Luque, G. Cortès, C. Segura, A. Maravilla, J. Esteban, and J. Fabregat, "End-to-end photopleth YsmographY (PPG) based biometric authentication by using convolutional neural networks," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 538–542.

[26] N. Karimian, M. Tehranipoor, and D. Forte, "Non-fiducial PPG-based authentication for healthcare application," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Feb. 2017, pp. 429–432.

[27] A. Sarkar, A. L. Abbott, and Z. Doerzaph, "Biometric authentication using photoplethysmography signals," in *Proc. IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2016, pp. 1–7.

[28] T. Choudhary and M. S. Manikandan, "Robust photoplethysmographic (PPG) based biometric authentication for wireless body area networks and m-health applications," in *Proc. 22nd Nat. Conf. Commun. (NCC)*, Mar. 2016, pp. 1–6.

[29] L. Pu, P. J. Chacon, H.-C. Wu, and J.-W. Choi, "Novel robust photoplethysmogram-based authentication," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4675–4686, Mar. 2022.

[30] G. Wu, J. Wang, Y. Zhang, and S. Jiang, "A continuous identity authentication scheme based on physiological and behavioral characteristics," *Sensors*, vol. 18, no. 2, p. 179, Jan. 2018.

[31] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[33] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.

[34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[35] R. Donida Labati, V. Piuri, F. Rundo, F. Scotti, and C. Spampinato, "Biometric recognition of PPG cardiac signals using transformed spectrogram images," in *Proc. Int. Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2021, pp. 244–257.

[36] J. Yang, Y. Huang, F. Huang, and G. Yang, "Photoplethysmography biometric recognition model based on sparse softmax vector and k-nearest neighbor," *J. Electr. Comput. Eng.*, vol. 2020, pp. 1–9, Oct. 2020.

[37] J. Yang, Y. Huang, R. Zhang, F. Huang, Q. Meng, and S. Feng, "Study on PPG biometric recognition based on multifeature extraction and naive Bayes classifier," *Sci. Program.*, vol. 2021, pp. 1–12, May 2021.

[38] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, Jul. 2013.

[39] S. Cresci, R. Di Pietro, and M. Tesconi, "Semantically-aware statistical metrics via weighting kernels," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2019, pp. 51–60.

[40] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.

[41] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[42] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999.

[43] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 413–422.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12 no. 10, pp. 2825–2830, 2012.

[45] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1914–1923, Aug. 2017.

[46] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. A. Chen, "Neurokit2: A Python toolbox for neurophysiological signal processing," *Behav. Res. Methods*, vol. 2021, pp. 1–8, Jan. 2021.

[47] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," *PLoS ONE*, vol. 8, no. 10, pp. 1–11, 2013.

[48] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

HUSSEIN A. ALY received the B.S. degree in computer science from Qatar University, Doha, Qatar, in 2020, and the master's degree in cybersecurity from Hamad Bin Khalifa University, Doha, in 2022.

He was an Intern with the Qatar Computing Research Institute, where he developed the backend service for the AIDER crisis response website. He is currently a Research Assistant with the Kindi Center for Computing Research. His current research interests include the application of artificial intelligence in biomedical engineering, privacy-preserving machine learning, and the application of artificial intelligence in smart grids.

ROBERTO DI PIETRO (Fellow, IEEE) received the M.S. degree in computer science and the M.S. degree in informatics from the University of Pisa, Italy, and the Post-M.S. Specialization Diploma degree in operations research and strategic decisions and the Ph.D. degree in computer science from the University of Rome "La Sapienza," Italy.

He has been working in the security field for more than 25 years, leading both technology-oriented and research-focused teams in the private sector (NOKIA Bell Labs), government (MoD), academia (HBKU, UniPD, and UniRomaTre), and international organizations (United Nations HQ, EUROJUST, IAEA, and WIPO), and other than being actively involved in strategic consultancy. He is currently a Full Professor of computer science with the Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), and affiliated with the Resilient Computing and Cybersecurity Center (RC3). Other than being involved in M&A of start-up-and having founded one (exited), he has been managing several multimillion-dollar security projects, producing more than 280 scientific articles, and 16 patents and patents applications over the cited topics, has coauthored three books, edited one, and contributed to a few others. His current research interests include AI-driven cyber-security, security, and privacy for wired and wireless distributed systems (e.g., DeFi, blockchain technology, cloud, the IoT, and OSNs), virtualization security, applied cryptography, intrusion detection, and data science.

Dr. Pietro is an ACM Distinguished Scientist and a member of Academia Europaea. From 2011 to 2012, he was awarded the Chair of Excellence from the University Carlos III, Madrid. In 2020, he received the Jean-Claude Laprie Award for having significantly influenced the theory and practice of dependable computing. In 2022, he received the Individual Inventor Award from HBKU. He is consistently ranked among the 2% Top World Scientists (Stanford's list) since this ranking existed.

Open Access funding provided by 'Qatar National Library' within the CRUI CARE Agreement