Don't You Forget About Me: A Study on Long-Term Performance in ECG Biometrics

Gabriel Lopes[1], João Ribeiro Pinto[1,2][0000−0003−4956−5902], and Jaime S. Cardoso[1,2][0000−0002−3760−2473]

[1] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
[2] Centre for Telecommunications and Multimedia, INESC TEC, Porto, Portugal

Corresponding author: João Ribeiro Pinto (jtpinto@fe.up.pt)

Abstract. The performance of biometric systems is known to decay over time, eventually rendering them ineffective. Focused on ECG-based biometrics, this work aimed to study the permanence of these signals for biometric identification in state-of-the-art methods, and measure the effect of template update on their long-term performance. Ensuring realistic testing settings, four literature methods based on autocorrelation, autoencoders, and discrete wavelet and cosine transforms, were evaluated with or without template update, using Holter signals from THEW's E-HOL 24h database. The results reveal ECG signals are unreliable for long-term biometric applications, and template update techniques offer considerable improvements over the state-of-the-art results. Nevertheless, further efforts are required to ensure long-term effectiveness in real applications.

Keywords: Biometrics · Electrocardiogram · Identification · Template Update

1 Introduction

The ability to identify or recognize another human being is of utmost importance. For access control and other purposes, security systems are based on unique credentials. Several methods are used such as PIN codes, passwords, ID cards, or keys [1]. These credentials are susceptible to be copied or counterfeit [2], and it is very easy for the person to forget, share, or lose them. The alternative is for systems to use something that is characteristic and belongs only to one person: a biometric trait [3]. Biometric traits currently include fingerprint, voice, face, and electrocardiogram (ECG), and only require the individual to be present when access is requested [4]. Relative to the traditional techniques, biometric systems are considered more secure, as traits are more difficult to counterfeited or steal than extrinsic credentials [5].

Current off-the-person ECG acquisition techniques, aiming towards increased simplicity, usability, and comfort, make ECG-based biometric systems effortless for the user [6]. However, when considering long-term usage, the performance

decays over time [7]. This applies to human-machine interfaces that require frequent identity control, especially those that are information-sensitive.

Variability on the input biometric data, biometric trait's aging, and variations caused by the subject's interaction with the sensor contribute to large intrasubject variability [5]. This makes stored individual templates to quickly lose representativity, resulting in poor recognition performance and placing serious challenges on long-term recognition. In biometrics in general, long-term identification requires frequent update of the templates to maintain acceptable performance over time [8], and avoid security flaws in information-sensitive systems.

Previously, Labati *et al.* [9] have studied the performance decay over time on their proposed algorithm for ECG-based authentication. In this paper, the goal was to build on their work, by studying the effect of ECG permanence and variability in biometric identification, and evaluating the impact of template update on long-term performance in four state-of-the-art algorithms for identification.

2   State-of-the-art review

In the literature, it is difficult to find a strong and widely accepted rule for template update. Most methods are based on heuristics and empirically determined thresholds, which are highly dependent on the data and application settings. For example, Komeili *et al.* [7], for authentication, have set the acceptance threshold equal to the point of zero false acceptance rate, thus ensuring updates with only genuine samples.

Nevertheless, it is possible to identify some common mechanisms that may vary depending on different factors: these include the choice of the update criterion (based on thresholds or graphs), the update periodicity (online or offline), the selection mechanism, and the template update working mode system (supervised or semi-supervised). The taxonomy of template update (see Fig. 1) divides the existing techniques in two categories: supervised and semi-supervised.

Supervised methods are offline methods in which label attribution is given by a supervisor. These contain the *Clustering* subcategory, which includes the *MDIST* that aims to search for the templates that minimize the intra distance among all the samples in the database (*i.e.*, the most similar) and *DEND* that aims to search for the templates that exhibit large intra-class variations resort to the dendrogram (*i.e.*, the most different) [11]. The second subcategory comprises Editing-based methods, which are independent of the number of templates and give focus on the whole collected training set T. A subset $E \in T$ is generated, maintaining the classification performance offered by T. The best subsets were obtained by reviewing the structure of the data (needs to be done to each subject) [8, 12]. All the algorithms (based on k-Nearest Neighbour) must be representative of T and can be roughly described as *incremental* when the E starts empty and grows, or *decremental* when E starts equal to T and in each iteration some instances are deleted, until some criterion is reached [8].
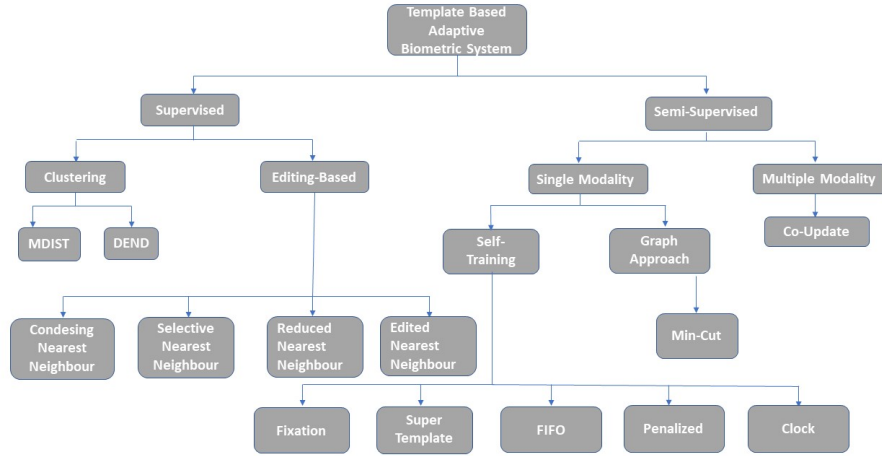
Fig. 1: Dendrogram representing the taxonomy of template update techniques (based on [10]).

Semi-supervised methods merge labelled (in biometrics represents the initial training samples) and unlabelled (correspond to the samples available during system operation) data to improve the system's performance. This category comprises the *Single Modality* (for unimodal biometric systems) and *Multiple Modality* (for systems using more than one biometric trait) subcategories. The *Single Modality* subcategory includes the *Self-Training* approaches such as *FIFO* (first-in-first-out), *Fixation*, *Super Template* (X composed by N templates x) where new genuine date is always fused to a common single template [13] updated online during the execution of continuous verification, *Penalized template update* method based on the mean of the past ECG's and the actual ECG [14] and *clock method* where the current template is tested against all the others stored in the database [15].

Generally, a new, unknown trait measurement is used for template update if its score (returned by the biometric recognition system) is above a set threshold. Hence, the future performance of the system relies heavily on the chosen threshold value [10].

The update threshold is commonly estimated using enrollment templates or training data. When training data are scarce or when using short enrollments, this leads to some problems: as important intrasubject variability is missed because use only the patterns similar to the templates stored; online methods are dependent on the order of the sequence of input data; vulnerability to large intra-class variations; since the algorithm normally looks for the minimal cost (high score), it can be stuck in local maximum and always use highly confident data for updating.

Table 1: Graph-based template update methods and their respective loss and regularizer functions (based on [16]).

Method	Source	Loss	Regularizer
Min Cut	[17]	$\sum_{i \in L}(y_i - y_{i|L})^2$	$\frac{1}{2}\sum_{ij} w_{ij}(y_i - y_j)^2$
Gaussian Random Fields and harmonic Function	[18]	$\sum_{i \in L}(f_i - y_i)^2$	$f^T \Delta f$
Local and Global Consistency	[19]	$\sum_{i=1}^{n}(f_i - y_i)^2$	$D^{-\frac{1}{2}}\Delta D^{\frac{1}{2}}$
Tikhonov Regularization	[20]	$\frac{1}{K}\sum_{i}(f_i - y_i)^2$	$\gamma f^T S f$
Manifold Regularization	[21]	$\frac{1}{l}\sum_{i=1}^{l} V(x_i, Y_i, f)$	$\gamma_A\|f\|_k^2 + \gamma_I\|f\|_I^2$
Graph Kernel from the Spectrum of Laplacian	[22]	$\min \frac{1}{2}w^T W$	$\exp(-\frac{\sigma}{2}\lambda)$
Spectral Graph Transducer	[18]	$\min c(f - \gamma)^T C(f - \gamma)$	$f^T L f$
Local Learning Regularization	[23]	$\min \frac{1}{k}\sum_{i=1}^{k}(y_i - f_k(x_i))^2$	$\frac{\gamma}{k}\|f_k\|^2$

Semi-supervised methods also include *Graph* approaches. These commonly define a graph where the nodes are either labelled (the identity is known) or unlabelled (unknown identity) data, and the edges (which can have different weights) are the similarity between those samples [10, 16]. To be considered a graph-based semi-supervised method, it must estimate a function f, approximate the known Y on the labelled nodes, and include two terms to turn the graph smooth: a loss function and a regularizer. These two terms are what defines each approach (as can be seen in Table 1) [16], among which the most common in biometrics is min-cut graphs [10].

Considering the topic of template update is still to be adequately addressed on ECG biometrics, this work studies the effect of ECG permanence and variability in long-term identification performance. Furthermore, it aimed to evaluate the effect of template update techniques, on the performance of several state-of-the-art methods.

3   Methods

3.1   Implemented identification methods

To fully and objectively evaluate the effects of ECG variability on the performance of biometric algorithms, a study was conducted on four literature methods, described below.

Plataniotis *et al.* [24] proposed an ECG biometric recognition method using a non-fiducial approach. Signals are preprocessed using a bandpass filter $(0.5 - 40$ Hz), followed by feature extraction with autocorrelation (AC) and dimensional-

ity reduction using discrete cosine transform (DCT). The fifteen most relevant features were selected, and Euclidean distance was used for classification.

Tawfik *et al.* [25] used a bandpass filter $(1-40$ Hz) in the preprocessing phase. QRS complexes (most stable part of ECG) were cut from the signal using a 0.35 second window. The average ensemble QRS was computed and features were extracted using DCT technique (thirty most relevant features were selected). A multilayer perceptron (MLP) is used for classification.

Belgacem *et al.* [26] also preprocessed signals with a bandpass filter $(1-40$ Hz). The QRS complexes were located and cut from the signal, and the average QRS was computed. The feature extraction resorted to Discrete Wavelet Transform (DWT). From all DWT decomposition levels, only the most relevant were selected, and a Random Forest is used for classification. The authors used this technique for authentication.

Eduardo *et al.* [27] used a Finite Impulse Response $(5-20$ Hz) filter for pre-processing. Heartbeats were cut with a fixed length of $[-200, 400]$ ms around each R peak. Outliers were detected and removed using DMEAN ($\alpha = 0.5$ and $\beta = 1.5$, with Euclidean distance). For classification, k-nearest neighbours (kNN) was used with $k = 3$ and cosine distance.

Standard sample wise normalization was performed (following Eq. 1) for all methods except that of Eduardo *et al.* [27], which required $[-1, 1]$ *min-max* normalization (Eq. 2).

$$x_{normalized}[g] = \frac{x[g] - \overline{x[g]}}{\sigma(x[g])} \tag{1}$$

$$x_{normalized}[g] = 2\left(\frac{x[g] - \min(x[g])}{\max(x[g]) - \min(x[g])}\right) - 1 \tag{2}$$

3.2   Template update methods

FIFO (first-in-first-out) is the most common strategy and, computationally, is very light. Here, the database is updated using new samples whose score is above or below a set threshold (whether the score represents similarity or dissimilarity, respectively), or between two threshold values (discarding previously stored sample) [7, 28]. The score of a new sample can either be output by a classifier, or be a measure of distance or similarity between that sample and the stored templates [29].

In this work, the training data were used to search for threshold values. Among all training samples, 75% were used to train a model, which was used to obtain scores for the remaining data samples. Comparing the scores with several thresholds, the error at each threshold was analysed (Fig. 2) to find one that simultaneously maximizes true positives and minimizes false positives.

Fixation consists on fixing certain templates, allowing only the remaining stored samples to be updated [30]. In this work, 25, 50, or 75% of the enrollment
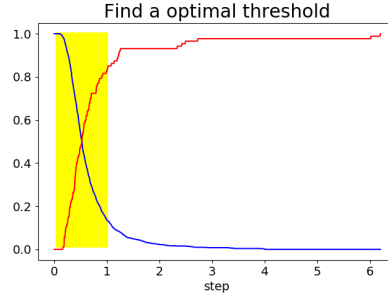
Find a optimal threshold

Fig. 2: Illustration of the search for the ideal threshold. The values were chosen near the intersection, inside the yellow zone.

templates of the individual are fixed, while the rest of the samples are free to be updated. This ensures some initial, labelled information of the subjects remains on the system over time.

An adaptation of this technique was explored. Here, $n + j \times n$ samples were fixed, where $n \in [1, 2, 3]$ is the number of fixed initial templates, and j increases over time. In this work, $j \in [0, 6]$ increased by one at each testing moment ($j \in [0, 6]$), which allowed the system to fix more and more samples over time, thus storing information on the subject's variability over time. In a real system with potentially endless use, the parameters n and j should be carefully chosen as to avoid the eventual fixation of the entire template gallery.

4   Experimental Settings

4.1   Dataset

For evaluation, the ECG signals used were from the E-HOL-03-0202-003 database[3] (most commonly designated as E-HOL 24h). This database consists of a study of 202 healthy subjects (only 201 were provided), recorded using three leads at 200 Hz sampling frequency, after an initial resting supine period of 20 minutes. From the available data of 201 subjects, thirteen were discarded due to saturation or unacceptable noise (subjects 1043, 9003, 9005, 9020, 9021, 9022, 9025, 9046, 9061, 9064, 9071, 9082 and 9105), similar to what was done by Labati *et al.* [13]. From each of the remaining 188 subjects, only the lead most closely resembling Lead I ECG was selected, to approximate off-the-person settings.

4.2   Experiments

In order to fit the used data, some changes were introduced on the original methods. On the method from Eduardo *et al.*, the cutoff frequencies of the

---

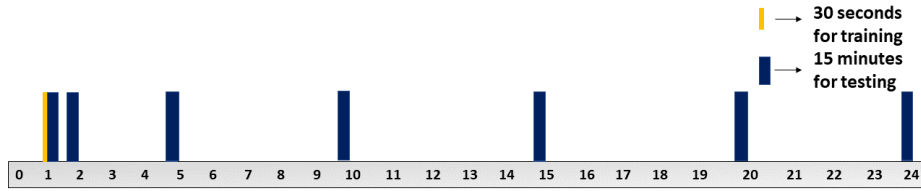[3] THEW. Available on: http://thew-project.org/Database/E-HOL-03-0202-003.html.

Fig. 3: Schema illustrating the use of each E-HOL records for training and testing (in orange - training segment; in blue - each test segment).

bandpass filter were changed to 1 and 40 Hz, to retrieve important information on higher frequencies; outlier removal was reparametrized with $\alpha = 1.2$ and $\beta = 1.5$. The autoencoder had the topology $[120, 60, 40, 20, 40, 60, 120]$ and was trained using the Adam optimizer with learning rate 0.01. Classification used $k = 1$.

For the method of Belgacem *et al.*, DWT feature extraction was performed in four decomposition levels, due to lower data sampling rate, and $cd4$, $cd3$, $cd2$, and $cd1$ coefficients were used as features.

Data were divided into train and test sets. The training phase used the last 30 seconds (mimicking short enrollments on real-life applications) of the first 60 minutes (avoiding unrealistic calm after the initial resting period) of each subject. Five-second overlap was used to obtain 26 samples from each 30 seconds of training data. To study performance over time, testing was performed over seven time points (see Fig. 3): one immediately after enrollment, another after one hour, and regularly until the end of the records. From each point, from 15 minutes of data, thirty 30 s samples are extracted, and batches are built with one sample from each of the 188 subjects.

5   Results

5.1   Without Template Update

After implementing, for identification, the method proposed by Labati *et al.* [9] (replicating their evaluation conditions), it was possible to conclude that the ECG signal is not fully permanent over 24 h. However, similarly to what was stated by Labati *et al.*, the results are relatively good over the first two hours (see Fig 4a), although permanence was not verified.

The performance results at each test hour, obtained through the weighted average of the corresponding batches, for the state-of-the-art methods can be found in Fig. 4b. It was found that the performance is mostly acceptable in the first test point, but performance decays significantly over time and variability changes considerably over the day.

Moreover, a minimum around the 15th hour occurs independently of the chosen method. Considering that most of the records start between 8-12 am,
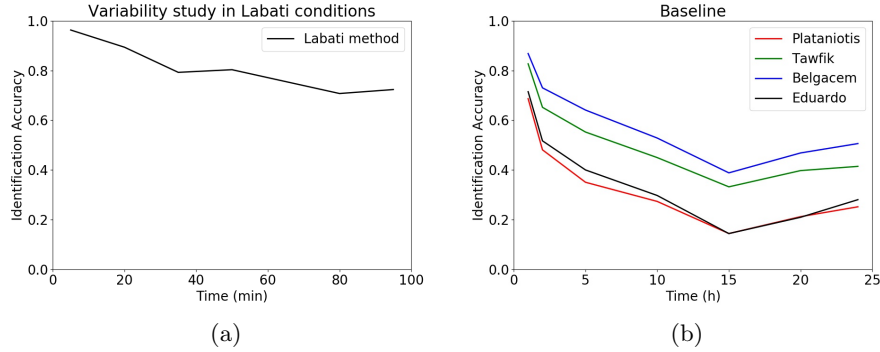
Fig. 4: Identification performance over time corresponding to (a) the Labati *et al.* method, and (b) the implemented state-of-the-art methods.

after 15 hours the subjects must be sleeping. In this perspective, it appears that the ECG is most different from normal when the subject is asleep.

5.2   With Template Update

Considering the previous results, template update was applied to the methods, in an effort to avoid performance decay over time. Fig. 5 presents the results using the FIFO technique, with diverse thresholds.

For the methods of Plataniotis *et al.* and Eduardo *et al.*, the best results were obtained using two thresholds, respectively, $\{0.3, 0.7\}$ (4.7% accuracy improvement) and $\{0.1, 0.3\}$ (+ 5.7% accuracy), improving all performance results until the 15th hour. However, for the method of Belgacem *et al.*, the performance worsens with template update after the first two hours (best results obtained when the difference between the highest and second highest scores $\Delta score \in [0.15, 0.3]$). The same was verified for the method of Tawfik *et al.* which, in the first two hours, offered best results with $\Delta score > 0.2$. In general, using two thresholds instead of one offered the best results.

Considering this, it appeared that the Random Forest and MLP classifiers are not suitable for these kind of template/model update. This was confirmed after a repetition of the evaluation of these methods, with kNN replacing the classifiers (see Fig. 6). With kNN, the template update was able to reduce the performance decay over time, improving accuracy, on average, by 7.9% and 9.2%, respectively, for the methods of Belgacem *et al.* and Tawfik *et al.*

As for the Fixation technique, the obtained results were more promising (see Fig. 7). This template update technique brought performance improvements for all methods. The fixation technique that offered the best results was $j \times 3 + 3$, improving the baseline identification accuracy, on average, by 10.0%.
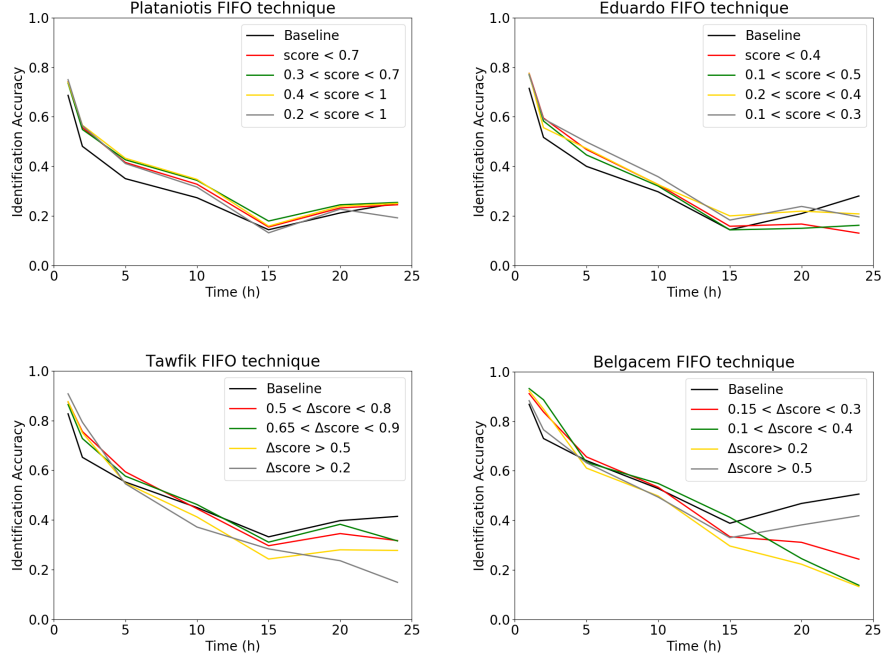
Fig. 5: Comparative FIFO methods with different thresholds using different identification methodologies.
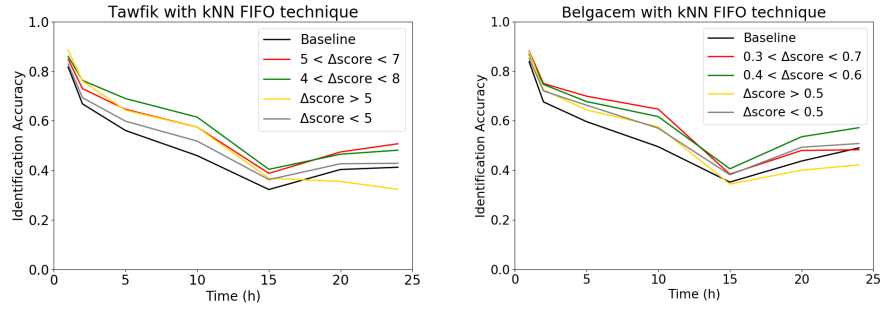

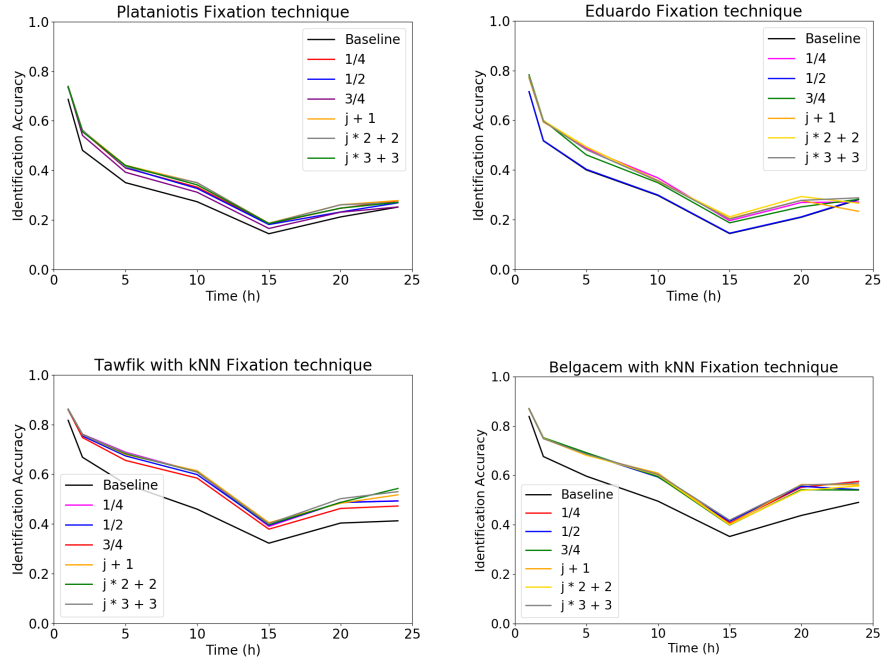Fig. 6: Results using FIFO update with different thresholds.

Fig. 7: Results using Fixation update (the corresponding value represents the number of samples that was fixated per subject).

6   Conclusion

This work studied how the ECG variability effects the performance of state-of-the-art biometric algorithms, and how template update could mitigate performance decay over time. The results have shown long-term identification performance in ECG biometrics is generally weak, despite the promising results often presented in the literature.

Template update techniques proved successful in enhancing the long-term performance of state-of-the-art methods, especially when using template fixation techniques. However, further efforts are needed for the study and development of more advanced techniques, with special focus on supervised techniques, so that ECG-based biometric systems can offer reliable performances over long periods of operation.

Acknowledgements

References

1. S. Prabhakar, S. Pankanti, and A.K. Jain. Biometric recognition: security and privacy concerns. *IEEE Security & Privacy Magazine*, 1(2):33–42, March 2003.
2. A. Hadid, N. Evans, S. Marcel, and Julian F. Biometrics Systems Under Spoofing Attack: An evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5):20–30, September 2015.
3. F. Agrafioti, F. M. Bui, and D. Hatzinakos. Secure Telemedicine: Biometrics for Remote and Continuous Patient Verification. *Journal of Computer Networks and Communications*, 2012:1–11, 2012.
4. Z. Akhtar, C. Micheloni, and G. Foresti. Biometric Liveness Detection: Challenges and Research Opportunities. *IEEE Security & Privacy*, 13(5):63–72, September 2015.
5. A.K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, January 2004.
6. J. R. Pinto, J. S. Cardoso, and A. Lourenco. Evolution, Current Challenges, and Future Possibilities in ECG Biometrics. *IEEE Access*, 6:34746–34776, 2018.
7. M. Komeili, N. Armanfard, and D. Hatzinakos. Liveness Detection and Automatic Template Updating Using Fusion of ECG and Fingerprint. *IEEE Transactions on Information Forensics and Security*, 13(7):1810–1822, July 2018.
8. B. Freni. *Template Editing and Replacement: novel methods for Biometric Template Selection and Update*. PhD thesis, University of Cagliari, Italy, 2010.
9. R. Labati, R. Sassi, and F. Scotti. ECG biometric recognition: Permanence analysis of QRS signals for 24 hours continuous authentication. In *2013 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 31–36, Guangzhou, China, November 2013. IEEE.
10. A. Rattani. *Adaptive Biometric System based on Template Update Procedures*. PhD thesis, University of Cagliari, 2010.
11. A. Lumini and L. Nanni. A clustering method for automatic biometric template selection. *Pattern Recognition*, 39(3):495–497, March 2006.
12. D. Hutchison and G. Anagnostopoulos. *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, SSPR & SPR 2008, Orlando, USA, December 4-6, 2008. Proceedings*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
13. R. D. Labati, V. Piuri, R. Sassi, F. Scotti, and G. Sforza. Adaptive ECG biometric recognition: a study on re-enrollment methods for QRS signals. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 30–37, Orlando, FL, USA, December 2014. IEEE.
14. S. Y. Chun. Small scale single pulse ECG-based authentication using GLRT that considers T wave shift and adaptive template update with prior information. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3043–3048, Cancun, Mexico, December 2016. IEEE.

15. T. Scheidat, A. Makrushin, and C. Vielhauer. Automatic Template Update Strategies for Biometrics. Technical report, Otto-von-Guericke University of Magdeburg, Germany, 2007.

16. X. Zhu. *Semi-Supervised Learning Literature Survey*. University of Wisconsin-Madison, USA, 2006.

17. A. Blum and S. Chawla. Learning from Labeled and Unlabeled Data using Graph Mincuts. In *Proceedings of International Conference on Machine Learning (ICML-2001)*, 2001.

18. X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. *Twentieth International Conference on Machine Learning (ICML-2003), Washington DC*, page 8, 2003.

19. Z. Zha, T. Mei, J. Wang, Z. Wang, and X. Hua. Graph-based semi-supervised learning with multiple labels. *Journal of Visual Communication and Image Representation*, 20(2):97–103, February 2009.

20. M. Belkin, I. Matveeva, and P. Niyogi. Regularization and Semi-supervised Learning on Large Graphs. In David Hutchison, editor, *Learning Theory*, volume 3120, pages 624–638. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

21. V. Sindhwani, P. Niyogi, M. Belkin, and S. Keerthi. Linear Manifold Regularization for Large Scale Semi-supervised Learning. page 4, August 2005.

22. O. Chapelle, B. Schölkopf, and A. Zien, editors. *Semi-supervised learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006.

23. Y. Kokkinos and K. Margaritis. Local learning regularization networks for localized regression. *Neural Computing and Applications*, 28(6):1309–1328, June 2017.

24. K. Plataniotis, D. Hatzinakos, and J. Lee. ECG Biometric Recognition Without Fiducial Detection. In *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pages 1–6, Baltimore, MD, USA, September 2006. IEEE.

25. M. M. Tawfik, H. Selim, and T. Kamal. Human identification using time normalized QT signal and the QRS complex of the ECG. In *2010 7th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP 2010)*, pages 755–759, July 2010.

26. N. Belgacem, A. Nait-Ali, R. Fournier, and F. Bereksi-Reguig. ECG Based Human Authentication using Wavelets and Random Forests. *International Journal on Cryptography and Information Security*, 2(2):1–11, June 2012.

27. A. Eduardo, H. Aidos, and A. Fred. ECG-based Biometrics using a Deep Autoencoder for Feature Learning - An Empirical Study on Transferability:. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, pages 463–470, Porto, Portugal, 2017. SCITEPRESS - Science and Technology Publications.

28. D. P. Coutinho, A. L. N. Fred, and M.A.T. Figueiredo. ECG-Based continuous authentication system using adaptive string matching:. In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 354–359, Rome, Italy, 2011. SciTePress - Science and and Technology Publications.

29. André Lourenço, Hugo Silva, and Ana Fred. Unveiling the Biometric Potential of Finger-Based ECG Signals. *Computational Intelligence and Neuroscience*, 2011:1–8, 2011.

30. J. Guerra-Casanova, C. Sánchez-Ávila, A. de Santos Sierra, and G. Bailador del Pozo. Score optimization and template updating in a biometric technique for authentication in mobiles based on gestures. *Journal of Systems and Software*, 84(11):2013–2021, November 2011.