This document will be distributed in print by Gesellschaft für Informatik through GI-Edition Lecture Notes in Informatics volume BIOSIG 2020, LNI volume #306, 978-3-88579-700-5.

A. Brömme, C. Busch, A. Dantcheva, K. Raja, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2020 1

Explaining ECG Biometrics: Is It All In The QRS?

João Ribeiro Pinto[1] and Jaime S. Cardoso[1]

Abstract: The literature seems to indicate that the QRS complex is the most important component of the electrocardiogram (ECG) for biometrics. To verify this claim, we use interpretability tools to explain how a convolutional neural network uses ECG signals to identify people, using on-the-person (PTB) and off-the-person (UofTDB) signals. While the QRS complex appears indeed to be a key feature on ECG biometrics, especially with cleaner signals, results indicate that, for larger populations in off-the-person settings, the QRS shares relevance with other heartbeat components, which it is essential to locate. These insights indicate that avoiding excessive focus on the QRS complex, using decision explanations during training, could be useful for model regularisation.

Keywords: Biometrics, Electrocardiogram, Explainability, Identification, Interpretability

1 Introduction

Throughout the past twenty years, research on biometrics based in the electrocardiogram (ECG) has largely been a success story [PCL18]. After successful proofs-of-concept in cleaner medical signals (*on-the-person*), the focus is quickly shifting to acquisitions in more realistic scenarios (*off-the-person*). Deep learning approaches [La18, Lu18, PCL19, PC19, Ha20] have been essential in dealing with the increased noise and variability in off-the-person settings, despite the performance and robustness issues that still hinder application in real scenarios.

However, deep learning decisions are obscure: unlike traditional methods based on fiducial features, we don't know what information the model uses to distinguish people. One can assume that the models look mainly to the QRS, since it is the most stable part of the ECG in the face of noise and variability [Sc00, HUvO01]. Several methods have thus focused on QRS complexes for ECG biometrics [Wa16, La18], but this practice has become uncommon in recent works. This indicates the true role of this waveform complex in identity discrimination is still to be adequately recognised.

Currently, pattern recognition researchers understand the importance of knowing what specific information is relevant for their models to reach decisions. Retreating to easily explainable traditional models (such as decision trees) is often unacceptable due to their performance limitations. Hence, various interpretability tools are being developed to peek into the inner workings of deep networks applied to diverse tasks [CPC19, SFC19, Se20].

This work uses, for the first time in the literature, such interpretability tools on a deep ECG biometric model, to understand what parts of the ECG are most useful for automatic human

[1] INESC TEC and Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, joao.t.pinto@inesctec.pt

identification. The model is a competitive state-of-the-art method [PCL19, PC19] applied for ECG-based identification in data subsets with diverse signal quality and number of identities. With this, we aim to assert the importance of the QRS and other waveforms for ECG biometrics and discuss future possibilities as this topic evolves towards more challenging and realistic scenarios. Additionally, we propose an intuitive way to visualise interpretations for unidimensional signals. The code and additional results are available online[2].

Besides this introduction, this paper presents some fundamental concepts on the ECG as a biometric trait, in section 2. The biometric identification model, the interpretability tools, and the visualisation method are described in section 3, and the experimental settings are detailed in section 4. Section 5 presents the obtained results and their discussion, and section 6 states the conclusions drawn from this work.

2 The Electrocardiogram as a Biometric Trait

The heart is composed of a muscle, the myocardium, that is responsible for its contraction and allows it to fulfil its purpose of pumping blood throughout the body [Ta09]. The myocardium contracts in response to depolarisation phenomena started by the atrioventricular node located on the interatrial septum. The waves of depolarisation that spread precisely across the heart are small electrical currents that can be measured using electrodes, resulting in the electrocardiogram (ECG) [MH13, Ta09].

Since the operation of the heart is a repetition of a sequence of phenomena, the ECG is approximately a cyclical repetition of a set of waveforms (P, Q, R, S, and T) that corresponds to a heartbeat (see Fig. 1) [MH13, PCL18]. The P wave is the first waveform and corresponds to the depolarisation of the myocardium cells in the atria. The Q, R, and S waveforms are commonly jointly considered as the QRS complex, which corresponds to the repolarisation of the atria and the depolarisation of the ventricles. The T wave corresponds to the repolarisation of the ventricles. This last wave is in some cases followed by a shorter waveform, the U wave, whose causes are still unclear [Ri08].

As a measurement of the electrical currents spread across the heart, the ECG signals will reflect the geometry of this organ. For example, larger hearts, with more cells to depolarise and repolarise, will result in ECG waveforms with larger amplitudes. Higher or lower basal heart rates will also result in different signal morphologies. Since heart geometry and basal heart rates vary across individuals, this intersubject variability is what makes the ECG sufficiently unique to be used in biometric recognition [HUvO01, vOHU00].

However, the ECG signals are also susceptible to intrasubject variability factors. Noise sources during acquisition, the short-term and long-term effects of exercise, emotional states, stress, drowsiness, and fatigue are some of the factors that reflect mainly in the heart rate variability, changing the morphology of the P-R and S-T segments [Sc00, ABH12]. These are the sources of uncertainty that hinder the use of the ECG as a biometric trait.

[2] xECG Repository. Available on: `https://github.com/jtrpinto/xECG`.

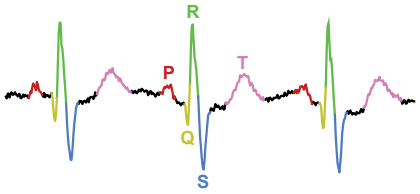

Fig. 1: Illustration of the ECG waveforms on a sample PTB signal segment.

While these are largely controlled on medical or on-the-person settings (where the subject is at rest, laying down, and signals are acquired using several high-quality gel electrodes), their effects are dominant for realistic off-the-person signals (acquired using few dry electrodes on the hands, during common daily activities) [Pi17, PCL18, PC19].

When compared with the P and T waves, the QRS corresponds to a larger polarisation event over a shorter period. In practice, this makes the QRS more dominant over noise and intrasubject variability than the other ECG waveforms [Pi17, PCL18]. Hence, the QRS is considered more stable over time and across variable conditions, which makes it better suited for biometric recognition.

Despite this, it is still unclear how much identity information is carried by the QRS complex compared to the other waveforms, and whether it is enough for an accurate and robust biometric recognition system. Studies on ECG-based biometric identification have shown it is possible to distinguish small sets of individuals in on-the-person settings using only the QRS complex or QRS fiducial amplitude and time measurements [Wa16, La18]. Nevertheless, this practice is becoming uncommon as research evolves towards realistic off-the-person signals and larger databases.

This denotes that the sole use of the QRS may not be adequate for off-the-person settings, or the individual information carried by the QRS may not be enough to distinguish individuals in large populations. This work aimed to address these doubts through a study on the role and relevance of the QRS and the other waveforms on ECG-based biometric identification. Interpretability tools are used to assess which parts of the ECG are more relevant to the decisions of an end-to-end identification model [PCL19], with on-the-person and off-the-person signals and data subsets with a varying number of identities.

3 Methodology

3.1 Biometric Identification Model

The biometric model for identification followed the architecture proposed by Pinto *et al.* [PCL19], which has attained state-of-the-art results in off-the-person settings for both identification and, later, authentication [PC19]. The model (see Fig. 2) receives five-second blindly segmented ECG signals and outputs probabilities for each of the N identities considered. Finding the highest probability score allows us to assign the respective identity to the input signal.

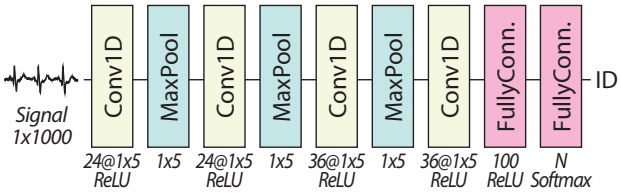

Fig. 2: Architecture of the biometric identification model

The model consists of an end-to-end 1D convolutional neural network (CNN) with four convolutional layers (with 1×5 filters, two layers with 24 followed by two with 36), followed by ReLU activation. Neighbouring convolutional layers are separated by 1×5 max-pooling layers. The last convolutional layer is followed by two fully-connected layers (100 neurons with ReLU and N neurons with softmax activation).

3.2 Interpretability Tools

To capture the dynamics behind the decisions of the biometric model, four interpretability methods are applied to the trained model: Occlusion, Saliency, Gradient SHAP, and DeepLIFT. Occlusion and Saliency are two of the simplest interpretability methods, while Gradient SHAP and DeepLIFT are more sophisticated and powerful. These are implemented in the Captum toolbox [Ko19] for PyTorch and are described below.

Occlusion The Occlusion method [ZF14] consists in measuring the influence of hiding a portion of the input on the output of the model. When hidden, the more relevant input parts will cause larger changes in the output, and will thus be assigned greater relevance in the explanations offered by this method. This is the simplest method to interpret a model, although the size of occluded regions should be carefully defined to obtain meaningful explanations.

Saliency The Saliency method [SVZ14] is based on the gradients of a model given a certain input. Through backpropagation, the gradient of target class scores w.r.t. the input is obtained. A saliency map is then generated by rearranging the class score derivatives, generating saliency maps that assign higher relevance to input regions that correspond to higher gradients. Requiring a single backpropagation pass, this method is a simple and fast way to obtain explanations on model predictions.

Gradient SHAP Gradient SHAP [LL17] is an approach based on game theory which considers the explanations of a model's predictions as models themselves. For sophisticated deep learning models, the explanation models are simplified and interpretable approximations of the respective models. SHapley Addictive exPlanation (SHAP) values, inspired by game theory's Shapley values, are computed through the gradient of a random

point between a baseline and the input with added random noise. The SHAP values denote how much a given part of the input raises the probability for the considered class, and are reportedly better aligned with human intuition and effective in discriminating among output classes.

DeepLIFT DeepLIFT (Deep Learning Important FeaTures) [SGK17] performs back-propagation to track the contributions to the output to the responsible parts of the input. Throughout this process, it compares the difference in inputs and outputs considering a reference (or baseline) input, assigning contribution scores to each neuron of the model. It also allows for the study of negative contributions: how much a specific part of the input contributes to lower the probability for the considered class.

3.3 Visualisation

Decision explanations obtained using interpretability tools are visualised using the multi-coloured line plot feature of Matplotlib [Hu07]. ECG signals are plotted so that the colour of each signal component represents its relative relevance for the decision. In this case, lighter yellow colours represent less relevant time samples, whereas more relevant samples assume darker purple colours. This way, both the ECG morphology and the relevance of each of its components are easily and intuitively presented.

4 Experimental Settings

The data used for model training and evaluation have been drawn from the Physikalisch-Technische Bundesanstalt ECG Database (PTB) [BKS95, Go00] and the University of Toronto ECG Database (UofTDB) [Wa14]. The PTB database includes on-the-person (high-quality) 12-lead ECG signals acquired at 1 kHz from 290 subjects at rest. The UofTDB includes single-lead off-the-person (more noisy and realistic) data acquired from 1019 subjects. To match the UofTDB, PTB signals were downsampled to 200 Hz and only Lead I was used.

Five-second segments were blindly extracted (without fiducial detection) from the recordings. Fifty per cent of those segments (*per* identity) were used during training and the remaining were reserved for testing. This provided more challenging test settings than those commonly found in the literature, but also deliberately avoided the most realistic settings (see [PC19]), for the sake of obtaining meaningful interpretations.

To simulate gradually increasing identification difficulty within each database, subsets of N identities are considered, with $N \in \{2, 5, 10, 20, 50, 100, 200, 500, 1019\}$. The identities in each subset are the first N in lexicographical order. Each subset includes all identities that compose smaller subsets, so subjects #1 and #2 are the main focus of analysis since these are present in all subsets. Throughout this paper, T_N denotes the subset of UofTDB data from N subjects and P_N denotes the subset of PTB data from N identities. As stated in

Tab. 1: True positive identification rate results (%) on the test data.

Database	Number of Identities								
	2	5	10	20	50	100	200[1]	500	1019
PTB	100.0	100.0	99.63	99.50	98.92	98.76	97.73	-	-
UofTDB	100.0	97.26	98.30	95.46	93.86	91.16	89.70	91.20	91.45

[1]For PTB, this column corresponds to the entire set of 290 subjects.

Table 1, P_{290} was used instead of P_{200} to take advantage of the entire PTB dataset. Model training details can be found online at this project's repository.

Performance evaluation is based on the True Positive Identification Rate (or accuracy): the fraction of test samples that are correctly assigned to their true identity by the trained model. Interpretations are examined through the proposed visualisation method.

5 Results and Discussion

The results of the performance evaluation are presented in Table 1. These results roughly follow the expected patterns considering the use of on-the-person *versus* off-the-person ECG data. The model is able to attain high true positive identification rates in both databases when the population is small, but as the set of subjects grows, performance decreases and a wide gap distinguishes the more challenging off-the-person settings from the more controlled on-the-person settings.

Additionally, one can find some unusual patterns in the performance results. Considering $M > N$, one would expect identification performance with subset T_N to be higher than with subset T_M. With UofTDB off-the-person data this is not always verified: *e.g.*, from T_5 to T_{10}, performance increases from 97.26% to 98.10%. In these cases, we need to consider that datasets with fewer identities have fewer data and, thus, more unstable results. Alternatively, the identities added to T_N to create T_M may be easier to discriminate ("sheep", according to the concept of biometric menagerie [Do98, YD10]) and thus contribute to improve accuracy. However, one should also regard the substantial regularisation needed to avoid overfitting and the instability during training as possible causes for these discrepancies. This is a very important insight into the increased difficulties of using off-the-person data and the need for improved and more robust biometric models.

Analysing the explanations obtained using the four interpretability tools (examples in Fig. 3 and Fig. 4), a trend is verified from smaller to larger identity subsets, consisting on the deviation from focusing mainly on the QRS complex to the increasing relevance of other parts of the heartbeats. This is also confirmed when combining the explanations of all heartbeats of each person into a single average heartbeat (see Fig. 5 and Fig. 6).

With the cleaner medical signals from PTB, the focus is mostly on the QRS complex, but information from other waveforms starts to become more and more relevant as more identities are added. It is noteworthy how, when discriminating PTB subjects #1 and #2 in a two-subject scenario (see Fig. 5), the model still focuses mainly on the QRS, even

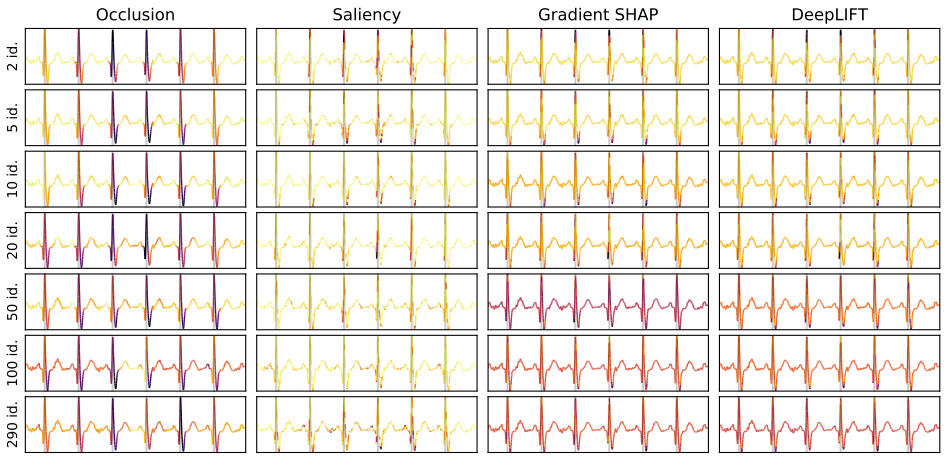

Fig. 3: Explanations over an example five-second ECG segment from PTB. In each subplot, the yellow to dark purple colours correspond to increasing time sample relevance and vertical grey lines denote R-peak locations. Signals were filtered for easy visualisation.

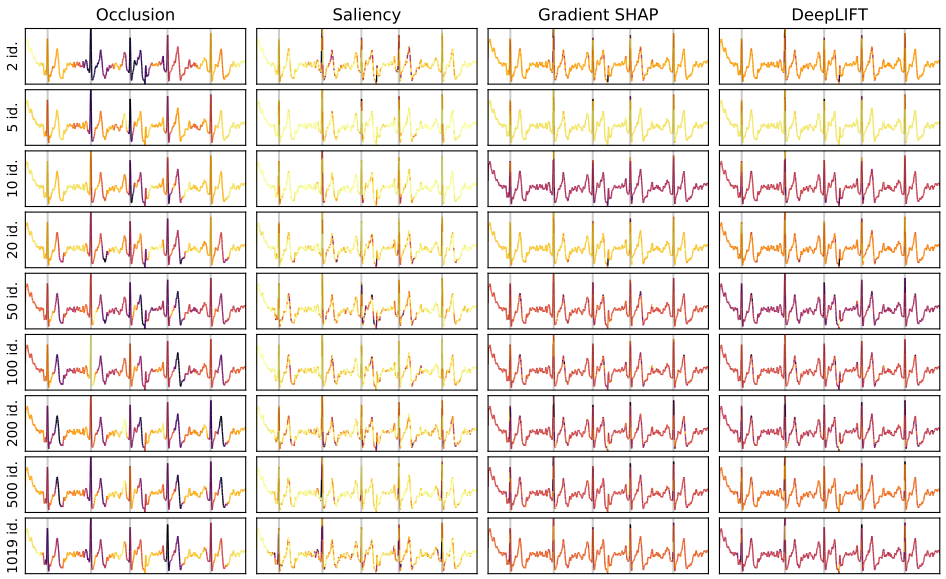

Fig. 4: Explanations over an example five-second ECG segment from UofTDB. In each subplot, the yellow to dark purple colours correspond to increasing time sample relevance and vertical grey lines denote R-peak locations. Signals were filtered for easy visualisation.

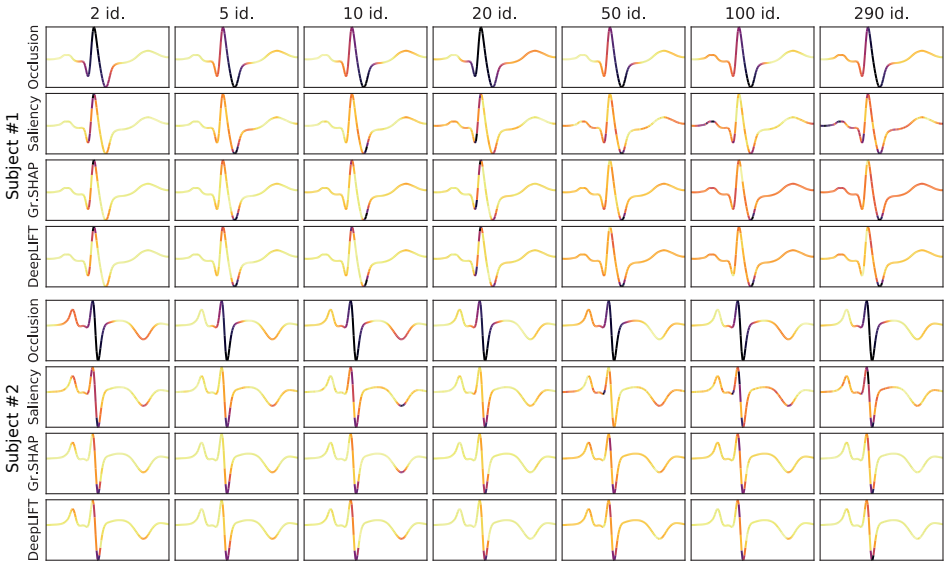

Fig. 5: Average explanations over heartbeat waveforms of subjects #1 (top) and #2 (bottom) on the subsets of the PTB database. In each subplot, the yellow to dark purple colours correspond to increasing time sample relevance. Signals were filtered for easy visualisation.

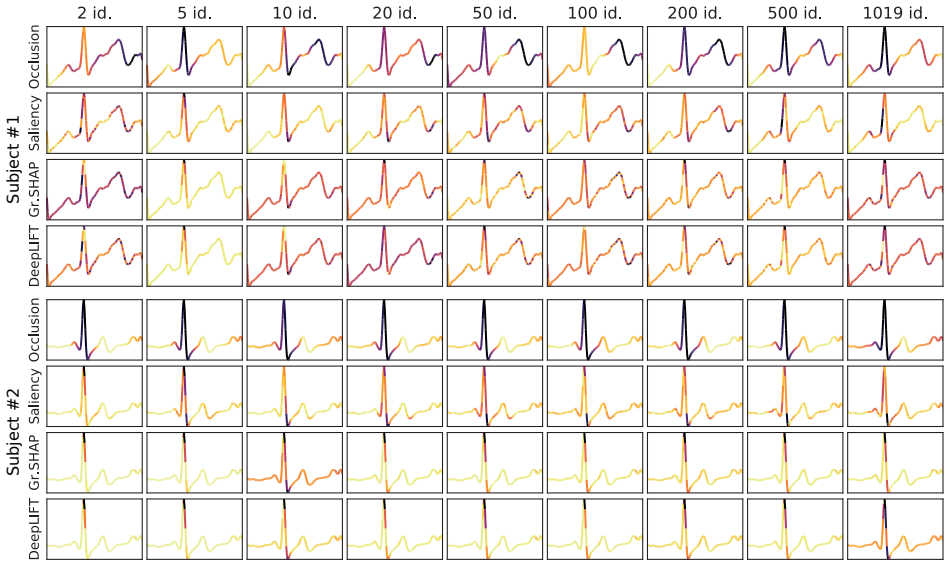

Fig. 6: Average explanations over heartbeat waveforms of subjects #1 (top) and #2 (bottom) on the subsets of the UofTDB database. In each subplot, the yellow to dark purple colours correspond to increasing time sample relevance. Signals were filtered for easy visualisation.

though subject #2 has a very specific characteristic, the inverted T-wave, that is arguably their most distinctive feature. This denotes how, in these cleaner signals, the QRS complex is so stable that the remaining waveforms, more susceptible to heart rate variability, are largely ignored by the model regardless of any visually obvious intersubject differences they may present.

With the more realistic off-the-person signals from UofTDB, the QRS retains high importance but the relevance is more evenly spread among the signal waveforms. In the specific case of subject #2 (see Fig. 6), it is evident that the QRS retains the highest importance for the decision, even in T_{1019} (the largest subset). This may denote that, even in these more challenging settings, the identification models will still give preference to the QRS over other waveforms if it is sufficiently unique among the considered identities. Nevertheless, in such large sets of identities, the expected behaviour is that of subject #1 (see Fig. 6), since the limited identity information carried by the QRS will lead the model to also look to other parts of the signal.

One interesting aspect is the difference between the results with Occlusion *versus* the other methods. Occlusion generally grants the QRS complex much more relevance, regardless of the settings. In the state-of-the-art approaches, the QRS complex is not only a source for identity features but also frequently used as an easily detectable reference landmark for the location of other ECG waveforms. This may also be the case in this end-to-end deep model. Although there are challenging contexts where the QRS may not be the main contributor to the decision, it may be essential to the deep model as a reference landmark to locate other waveforms in the signal. Hence, when occluded, it will be the signal component that most impacts the decision, causing the occlusion method to generally consider it the most relevant.

6 Conclusion

This work aimed to explain how deep models use ECG signals to distinguish people, using interpretability tools. Overall, the obtained results partially confirm the claim that the QRS is the key to ECG-based biometrics. With small populations in on-the-person settings, it can alone be used for reliable recognition. However, as we evolve towards larger populations and off-the-person settings, other components become relevant in discriminating people, as the models require more identity information to overcome the hurdles placed by enhanced intrasubject variability.

However, even though relevance is more evenly shared in off-the-person identification in large sets of identities, the QRS is shown as essential by the occlusion method. It appears that, just like several literature methods, the implemented end-to-end model learnt to use the QRS as a landmark for the location of other ECG components in the signal, resulting in large output changes when the QRS is occluded. Hence, despite the literature claims, one should avoid relying too heavily on any single part of the ECG, including the QRS complex, since all waveforms carry identity information that proves increasingly useful in more realistic settings and larger populations.

Beyond these insights, further efforts should be devoted to extend this study and offer a deeper, more thorough, and more objective analysis of the contribution of each ECG waveform to the model's decisions. Obtaining more systematic and complete explanations could create new opportunities on the use of interpretability tools during model training. Using explanations to regularise models and promote focus in the most relevant signal components or the distributed use of the whole signal (instead of just the QRS) could lead to improved recognition accuracy and robustness.

Acknowledgements

This work was financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalization - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project "POCI-01-0145-FEDER-030707", and within the PhD grant "SFRH/BD/137720/2018". The authors wish to thank the creators and administrators of the PTB (Physikalisch-Technische Bundesanstalt, Germany) and UofTDB (University of Toronto, Canada) databases, which have been essential for this work.

References

[ABH12] Agrafioti, F.; Bui, F. M.; Hatzinakos, D.: Secure Telemedicine: Biometrics for Remote and Continuous Patient Verification. Journal of Computer Networks and Communications, 2012:11, 2012.

[BKS95] Bousseljot, R.; Kreiseler, D.; Schnabel, A.: Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, 40(1), 1995.

[CPC19] Carvalho, D. V.; Pereira, E. M.; Cardoso, J. S.: Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics, 2019.

[Do98] Doddington, G.; Liggett, W.; Martin, A.; Przybocki, M.; Reynolds, D.: Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. Technical report, National Institute of Standards and Technology, 1998.

[Go00] Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, J.; Ivanov, P. C.; Mark, R.; Stanley, H. E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation, 101(23):e215–e220, 2000.

[Ha20] Hammad, M.; Pławiak, P.; Wang, K.; Acharya, U. R.: ResNet-Attention model for human authentication using ECG signals. Expert Systems, p. e12547, 2020.

[Hu07] Hunter, J. D.: Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3):90–95, 2007.

[HUvO01] Hoekema, R.; Uijen, G. J. H.; van Oosterom, A.: Geometrical aspects of the interindividual variability of multilead ECG recordings. IEEE Transactions on Biomedical Engineering, 48(5):551–559, May 2001.

[Ko19] Kokhlikyan, N.; Miglani, V.; Martin, M.; Wang, E.; Reynolds, J.; Melnikov, A.; Lunova, N.; Reblitz-Richardson, O.: , PyTorch Captum. https://github.com/pytorch/captum, 2019.

[La18] Labati, R. D.; Muñoz, E.; Piuri, V.; Sassi, R.; Scotti, F.: Deep-ECG: Convolutional Neural Networks for ECG biometric recognition. Pattern Recognition Letters, 126:78–85, 2018.

[LL17] Lundberg, S. M.; Lee, S.-I.: A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems. pp. 4765–4774, 2017.

[Lu18] Luz, E. J. S.; Moreira, G. J. P.; Oliveira, L. S.; Schwartz, W. R.; Menotti, D.: Learning Deep Off-the-Person Heart Biometrics Representations. IEEE Transactions on Information Forensics and Security, 13(5):1258–1270, May 2018.

[MH13] Marieb, E. N.; Hoehn, K.: Human Anatomy & Physiology. Pearson, Glenview, IL, ninth edition, 2013.

[PC19] Pinto, J. R.; Cardoso, J. S.: An End-to-End Convolutional Neural Network for ECG-Based Biometric Authentication. In: 10th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS). Tampa, FL, United States, 2019. (in press).

[PCL18] Pinto, J. R.; Cardoso, J. S.; Lourenço, A.: Evolution, Current Challenges, and Future Possibilities in ECG Biometrics. IEEE Access, 6:34746–34776, June 2018.

[PCL19] Pinto, J. R.; Cardoso, J. S.; Lourenço, A.: Deep Neural Networks For Biometric Identification Based On Non-Intrusive ECG Acquisitions. In: The Biometric Computing: Recognition and Registration, chapter 11, pp. 217–234. CRC Press, 2019.

[Pi17] Pinto, J. R.; Cardoso, J. S.; Lourenço, A.; Carreiras, C.: Towards a Continuous Biometric System Based on ECG Signals Acquired on the Steering Wheel. Sensors, 17(10), 2017.

[Ri08] Riera, A. R. P.; Ferreira, C.; Filho, C. F.; Ferreira, M.; Meneghini, A.; Uchida, A. H.; Schapachnik, E.; Dubner, S.; Zhang, L.: The enigmatic sixth wave of the electrocardiogram: the U wave. Cardiology Journal, 15(5):408–421, 2008.

[Sc00] Schijvenaars, R. J. A.: Intra-individual Variability of the Electrocardiogram: Assessment and exploitation in computerized ECG analysis. PhD thesis, Erasmus University Rotterdam, 2000.

[Se20] Sequeira, A. F.; Silva, W.; Pinto, J. R.; Goncalves, T.; Cardoso, J. S.: Interpretable Biometrics: Should We Rethink How Presentation Attack Detection is Evaluated? In: 2020 8th International Workshop on Biometrics and Forensics (IWBF). 2020.

[SFC19] Silva, W.; Fernandes, K.; Cardoso, J. S.: How to produce complementary explanations using an Ensemble Model. In: 2019 International Joint Conference on Neural Networks (IJCNN). 2019.

[SGK17] Shrikumar, A.; Greenside, P.; Kundaje, A.: Learning Important Features through Propagating Activation Differences. In: 34th International Conference on Machine Learning. pp. 3145–3153, 2017.

[SVZ14] Simonyan, K.; Vedaldi, A.; Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In: Workshop at International Conference on Learning Representations. 2014.

[Ta09] Tate, P.: Seeley's Principles of Anatomy and Physiology. McGraw-Hill, New York, NY, second edition, 2009.

[vOHU00] van Oosterom, A.; Hoekema, R.; Uijen, G. J. H.: Geometrical factors affecting the interindividual variability of the ECG and the VCG. Journal of Electrocardiology, 33:219 – 227, 2000.

[Wa14] Wahabi, S.; Pouryayevali, S.; Hari, S.; Hatzinakos, D.: On Evaluating ECG Biometric Systems: Session-Dependence and Body Posture. IEEE Transactions on Information Forensics and Security, 9(11):2002–2013, Nov 2014.

[Wa16] Waili, T.; Nor, R. M.; Rahman, A.; Sidek, K. A.; Ibrahim, A. A.: Electrocardiogram Identification: Use a Simple Set of Features in QRS Complex to Identify Individuals. In: 12th International Conference on Computing and Information Technology (IC2IT). pp. 139–148, 2016.

[YD10] Yager, N.; Dunstone, T.: The Biometric Menagerie. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(2):220–230, February 2010.

[ZF14] Zeiler, M. D.; Fergus, R.: Visualizing and Understanding Convolutional Networks. In: 2014 European Conference on Computer Vision (ECCV). pp. 818–833, 2014.