xECG: Using Interpretability to Understand Deep ECG Biometrics

João Ribeiro Pinto[1,2]
joao.t.pinto@inesctec.pt
Jaime S. Cardoso[1,2]
jaime.cardoso@inesctec.pt

INESC TEC
Porto, Portugal
Faculdade de Engenharia, Universidade do Porto
Porto, Portugal

Abstract

The literature on electrocardiogram (ECG) biometrics frequently indicates that the QRS complex is the most important part of this signal. Some go further and argue that just the QRS is enough for accurate ECG biometric identification. To verify this claim, this work uses interpretability tools to analyse how a state-of-the-art deep learning model uses each part of the ECG signal to reach identity decisions under different signal quality conditions and population sizes. Results indicate that the QRS is indeed the most relevant part of the ECG for identification, but its relative importance is smaller in more realistic scenarios. Such insights could be used as regularisation to avoid excessive focus on the QRS complex and thus achieve more robust models.

1 Introduction

The ECG measures the conduction of electrical potentials across the heart's muscle that controls its contraction and relaxation. The cyclical repetition of depolarisation and repolarisation of cells on the heart's atria and ventricles turns the ECG into a repetition of easily recognisable heartbeats. Each heartbeat is composed of a P wave, a QRS complex, and a T wave and, since its morphology depends on the heart structure, it carries important identity information [9].

ECG signals have been successfully used for several automatic pattern recognition tasks, including biometric recognition [9]. Once focused on clean medical signals (*on-the-person* signals), the field of ECG biometrics has decisively evolved towards signals acquired in more realistic biometric scenarios (*off-the-person*). In such scenarios, where noise and variability are prevalent, deep learning methods [3, 5, 7, 10] have enabled higher accuracy and robustness.

However, while traditional methods based on fiducial features are fairly transparent, it is not easy to understand what information deep learning models use to reach a decision. Considering the well-known stability and uniqueness of the QRS complex [4], one can assume that models will focus primarily on this landmark, but this assumption may be incorrect. While pioneer methods used only information from the QRS, this practice has become increasingly uncommon. This may indicate that the QRS may not be sufficient for recognising identity in some scenarios.

These doubts on the role played by the signal waveforms in modern ECG biometric recognition can be understood through interpretability tools. These tools have been developed following the growing awareness of the paramount importance of transparency in artificial intelligence. They enable the analysis of the inner workings of machine learning models applied to diverse pattern recognition tasks [2]. Hence, instead of returning to simpler algorithms such as decision trees (sacrificing performance in favour of transparency), interpretability allows us to obtain sophisticated models which are both highly accurate and transparent.

Hence, the work described in this paper[1] aims to understand the behaviour of deep ECG biometrics through interpretability [8]. A state-of-the-art model for ECG identification [7, 10] was trained on diverse scenarios, with on-the-person and off-the-person data from growing sets of identities, emulating increasingly challenging conditions. Then, interpretability tools were used to evaluate which parts of the ECG signal are most relevant to the model's decisions.

2 Methodology

This study consisted of (1) training an identification model in diverse scenarios, (2) inferring for test samples and analysing identification accuracy, (3) obtaining decision explanations for each test sample, and (4) visualising and analysing the results. Details on the methods are presented below.

[1]Code and additional results available at `https://github.com/jtrpinto/xECG`.

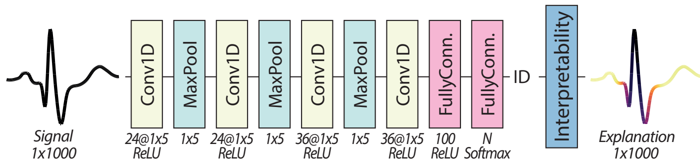

Figure 1: Schema of the methodology: the CNN model is trained to offer identity decisions, which are then explained by interpretability tools.

Model The model used for identification (see Fig. 1) is the one proposed by Pinto *et al.* [7, 10], which has achieved state-of-the-art performance in both identification and identity verification. It receives a five-second ECG segment as input and delivers probabilities for each identity on the training set. The model is an end-to-end convolutional neural network (CNN) with four convolutional layers (two with 24 1×5 filters followed by two with 36 filters) interposed with three 1×5 max-pooling layers. To deliver N identity probabilities, the model concludes with two fully-connected layers (100 and N neurons) and softmax activation.

Interpretability Tools To quantify the relative importance of ECG waveforms on the decisions of the trained identification models, four interpretability tools, from the Captum toolbox for PyTorch, were used:
- *Occlusion* [14] works by hiding parts of the input and measuring the corresponding change in the model's outputs. More relevant input regions will correspond to larger changes in the output;
- *Saliency* [12] uses backpropagation to compute the gradients of target class scores w.r.t. the input. The resulting class score derivatives are rearranged into a saliency map, which assigns higher relevance to input regions that correspond to higher gradients;
- *Gradient SHAP* [6] considers the explanations of a model's predictions as models themselves. Explanation models are simplified and interpretable approximations of the sophisticated models that generate them. SHapley Addictive exPlanation (SHAP) values are computed and denote how much each input region raises the probability for a given class;
- *DeepLIFT* (Deep Learning Important FeaTures) [11] uses backpropagation to trace output contributions to the responsible regions of the input. It compares differences in inputs and outputs based on a baseline input and assigns contribution scores to each neuron of the model.

Visualisation Just as in image interpretability, signal explanations should be visualised in a way that illustrates input morphology and local relevance simply and clearly. Thus, this work proposes a visualisation methodology based on multicoloured line plots. The colour of each part of the signal depends on its relevance for the model's decision: less relevant parts are in light yellow and most relevant parts are in darker purple.

3 Experimental Setup

This work used data from the PTB ECG database [1] and the University of Toronto ECG database (UofTDB) [13]. The PTB includes on-the-person signals from 290 subjects at rest. The UofTDB includes off-the-person recordings from 1019 volunteers during up to six sessions and five postures. Five-second segments were blindly extracted from the recordings. Half of the segments from each identity were used during training and the remaining were used for testing.

Growing subsets of identities (subjects) are considered to emulate increasing challenging scenarios. Within each database, the N first identities are selected, with $N \in \{2, 5, 10, 20, 50, 100, 200, 500, 1019\}$. Identities #1 and #2 are thus the only ones present in all subsets. To take full advantage of the PTB dataset, the entire set of 290 identities was used instead of the 200 identities subset.

Table 1: Identification accuracy 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.

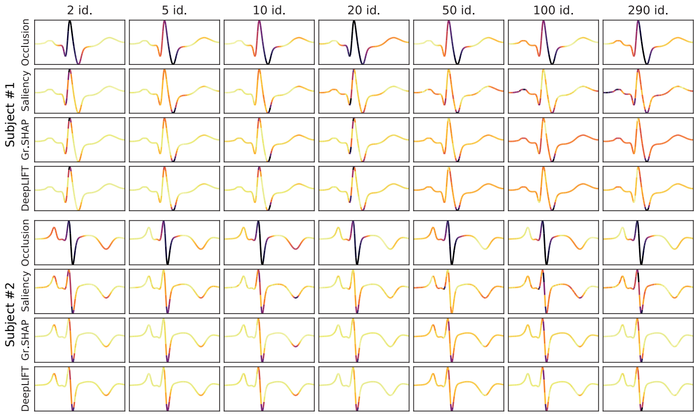

Figure 2: Model explanations on the PTB database.

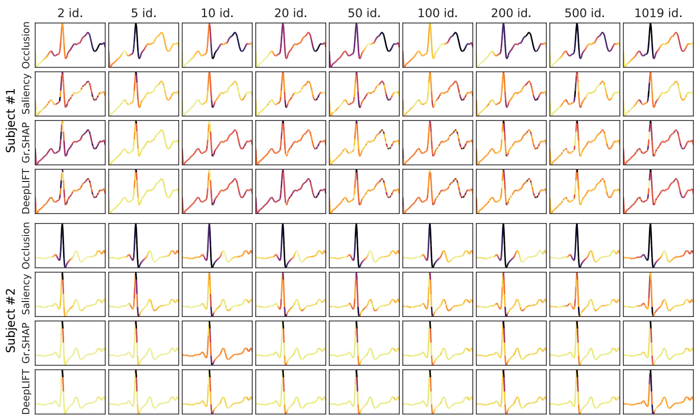

Figure 3: Model explanations on the UofTDB database.

4 Results and Discussion

After training in the aforementioned diverse scenarios, the models' identification accuracy was evaluated (see Table 1). The results follow the expected trends: the model is significantly more accurate in smaller populations for both databases. However, increasing the number of subjects leads to faster performance decay when using off-the-person data. This is evidence of the yet unsolved challenges of increased noise and variability in off-the-person ECG biometrics.

The explanations obtained using the interpretability tools were combined into average heartbeat relevance maps for each subject in each scenario (see Fig. 2 and Fig. 3). One can observe that, with cleaner signals from PTB, acquired in medical conditions, the models focus mainly on the QRS complex. However, as the set of identities grows, the models start to use some information about other waveforms for their decisions. Nevertheless, with these signals, the stability and information of the QRS complex are enough for the model to largely ignore the rest of the signal.

With more realistic signals from UofTDB, the QRS, although important, shares the relevance for the decision more evenly with the other ECG waveforms. These results show that, while the models still prefer the QRS complex, the challenging scenarios of noisy signals and larger populations lead them to look for broader sources of identity information and take advantage of the other ECG waveforms. Another interesting result is that Occlusion generally attributes greater relevance to the QRS than other interpretability tools. This occurs even in the most challenging scenarios. This shows that the deep learning model, not unlike traditional methods, may be learning to locate the different ECG waveforms using the QRS as a reference landmark: when occluded, the effect on the output is larger than with other interpretability methods.

5 Conclusion

This work used interpretability tools to study how state-of-the-art deep biometric models use ECG signals to distinguish people. In general, this study found that the literature is correct in hailing the QRS complex as the most important part of the ECG for biometrics. However, this is more evident in less challenging scenarios. Considering more realistic scenarios, with larger populations and lower quality signals, the relevance is more evenly shared between the ECG waveforms. This indicates that, although still important, the QRS is no longer enough for robust identification.

Thus, one should avoid placing the entire burden of identification upon any single part of the ECG, including the QRS. Every part of the signal carries information that is important in realistic scenarios. These insights could be used as regularisation to promote behaviours that lead towards more accurate and robust models. Additionally, further efforts should be devoted to extending this study into more thorough and objective analyses of ECG waveform relevance.

Acknowledgements

This work was partially funded by the Project TAMI - Transparent Artificial Medical Intelligence (NORTE-01-0247-FEDER-045905) financed by ERDF - European Regional Fund through the North Portugal Regional Operational Program - NORTE 2020 and by the Portuguese Foundation for Science and Technology - FCT under the CMU - Portugal International Partnership and the Ph.D. grant "SFRH/BD/137720/2018".

References

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

[2] D. Carvalho, E. Pereira, and J. S. Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 2019.

[3] M. Hammad, P. Pławiak, K. Wang, and U. R. Acharya. ResNet-Attention model for human authentication using ECG signals. *Expert Systems*, 2020.

[4] R. Hoekema, G. J. H. Uijen, and A. van Oosterom. Geometrical aspects of the interindividual variability of multilead ECG recordings. *IEEE-TBME*, 48(5):551–559, 2001.

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

[6] S. M. Lundberg and S.-I. Lee. A Unified Approach to Interpreting Model Predictions. In *NeurIPS*, pages 4765–4774, 2017.

[7] J. R. Pinto and J. S. Cardoso. An End-to-End Convolutional Neural Network for ECG-Based Biometric Authentication. In *BTAS*, 2019.

[8] J. R. Pinto and J. S. Cardoso. Explaining ECG Biometrics: Is It All In The QRS? In *BIOSIG*, 2020.

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

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

[11] A. Shrikumar, P. Greenside, and A. Kundaje. Learning Important Features through Propagating Activation Differences. In *ICML*, pages 3145–3153, 2017.

[12] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. In *ICLR*, 2014.

[13] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos. On Evaluating ECG Biometric Systems: Session-Dependence and Body Posture. *IEEE-TIFS*, 9(11):2002–2013, 2014.

[14] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. In *ECCV*, pages 818–833, 2014.