Optimisation of artefact detection in photoplethysmography heart rate data: influence of different classifiers in machine learning models

P. Vermunicht¹, C. Buyck¹, K. Makayed², S. Herman², J.S. Piedrahita Giraldo³, K. Laukens³, L. Knaepen⁴, S. Naessens⁵, W. Hens⁵, E. Van Craenenbroeck⁵, L. Desteghe¹, H. Heidbuchel⁵

¹University of Antwerp, Research Group Cardiovascular Diseases, Antwerp, Belgium ²University of Antwerp, Department of Computer Science, Antwerp, Belgium ³University of Antwerp, Biomedical Informatics Research Network Antwerp (Biomina), Antwerp, Belgium ⁴Hasselt University, Faculty of Medicine and Life Sciences, Hasselt, Belgium ⁵University Hospital Antwerp, Department of Cardiology, Antwerp, Belgium **Funding Acknowledgements:** Type of funding sources: Public grant(s) – National budget only. Main funding source(s): This project is supported by the Fund for Scientific Research, Flanders (G084023N)

Background: Heart rate (HR) tracking by wrist-worn devices using photoplethysmography (PPG) could assist in continuously following up physical activity. However, the accuracy can be impacted by (motion) artefacts. Machine learning models could help to recognise artefacts in PPG-based HR data. The choice of classifier in these machine learning models is a determing factor for task performance of the model.

Purpose: This study evaluates and determines the optimal classifier for a new machine learning-based approach to enhance the reliability of artefact detection in PPG-based HR data.

Methods: A total of 62 participants (27 cardiac rehabilitation patients, 35 healthy athletes) wore both a test device and a reference device measuring HR continuously for 24 hours. A training dataset was prepared, assigning two independent labels (i.e. anomaly and activity) to each HR episode based on the reference device data. Fitbit data were processed using our in-house designed artefact removal procedure, which involves the application of two classification models: one for anomaly detection and another for activity detection. Four distinct classifiers were employed for both models: Balanced Bagging, Balanced Bagging with Random Forest, Balanced Random Forest, and Logistic Regression. Each classifier was evaluated using area under the receiver operating characteristic curve (ROC-AUC), accuracy, sensitivity and specificity.

Results: Of the 1,647,328 HR data points collected, 103,095 (6.26%) were identified as artefacts. Figure 1 and Figure 2 summarise the performance of the distinct classifiers for the anomaly model and the activity model, respectively. Balanced Bagging and Balanced Bagging with Random Forest consistently demonstrate the highest AUC values and accuracies across both anomaly and activity detection models (anomaly detection: AUC = 0.95, accuracy = 89-85%; activity detection: AUC = 0.98, accuracy = 95%). Comparing these two, Balanced Bagging with Random Forest emerges as the preferred option, given the highest sensitivity in both anomaly detection (93%>86%) and activity detection models (99%>96%). In contrast, Balanced Random Forest and Logistic Regression exhibit inferior performance. In the anomaly detection model, Balanced Random Forest exhibits a lower sensitivity of 75%, while Logistic Regression performs even worse with a sensitivity of 25%. Similarly, in the activity detection model, both Balanced Random Forest and Logistic Regression demonstrate diminished performance.

Conclusions: Balanced Bagging with Random Forest emerges as the optimal classifier to detect anomalies and activities in continuous PPG-based HR data, thus contributing to the optimisation of our in-house designed procedure for removing artefacts. This processing aims to provide a reliable and automatic way for continuous HR monitoring, which can help monitor and guide physical activities.

Figure 1: Performance of the **anomaly detection model** with different classifiers: **(A)** Balanced Bagging, **(B)** Balanced Bagging with Random Forest, **(C)** Balanced Random Forest, and **(D)** Logistic Regression.

Receiver operating characteristic (ROC) curves with the corresponding area under the curve (AUC) values are displayed. The accuracy, sensitivity and specificity shown are the ones at a classification threshold of 0.5. Values over 80% are green, 60%-80% are orange, and below 60% are red.



Figure 2: Performance of the **activity detection model** with different classifiers: **(A)** Balanced Bagging, **(B)** Balanced Bagging with Random Forest, **(C)** Balanced Random Forest, and **(D)** Logistic Regression.

Receiver operating characteristic (ROC) curves with the corresponding area under the curve (AUC) values are displayed. The accuracy, sensitivity and specificity shown are the ones at a classification threshold of 0.5. Values over 80% are green, 60%-80% are orange, and below 60% are red.

