Factors for incorrect automated atrial fibrillation identification in large-scale screening

M.D. Zink¹, M. Lueken², S. Leonhardt², B. Freedman³, A.P. Keszei⁴, K. Mischke⁵, N. Marx⁶

¹RWTH University Hospital Aachen, Aachen, Germany
²RWTH Aachen University, Chair for Medical Information Technology (MedIT) Helmholtz-Institute for Biomedical En, Aachen, Germany
³Heart Research Institute, Sydney, Australia
⁴RWTH University Hospital Aachen, CTC-A, Aachen, Germany
⁵Leopoldina Hospital, Cardiology, Schweinfurt, Germany
⁶RWTH University Hospital Aachen, Internal Medicine I, Cardiology, Pulmonology & Vascular Medicine, Aachen, Germany

Funding Acknowledgements: Type of funding sources: Public grant(s) – EU funding. Main funding source(s): Pfizer/BMS and European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No 847770

Introduction: Correct heart rhythm diagnosis in single-lead electrocardiogram (ECG) screening for atrial fibrillation (AF) is paramount for large-scale implementation.

Aim: Examine the accuracy of automated AF detection and identify patient factors and ECG signal quality leading to incorrect measurements based on human overread and a machine learning approach.

Methods: Screening for AF using a single-lead ECG device (MyDiagnostick, Biomedical Systems, Netherlands) was performed on 7295 customers in community pharmacies in Germany. ECG rhythm strips were blinded for automatic heart rhythm analysis and were analyzed by human overread for heart rhythm and signal quality from excellent to uninterpretable (Figure 1A). A machine learning-based estimation was used to score signal quality (MLSQE). Inclusion criteria were 30 seconds of consecutive ECG recording and complete data of pharmacy visit. In total, 6732 participants were eligible for analysis.

Results: An incorrect automated heart rhythm diagnosis was found for 206 subjects (false positive: 112; false negative: 94). By human overread 80% (N=5377) of the recorded ECGs were rated with excellent and good signal quality in which heart rhythm diagnosis based on visible atrial electrical activity (Figure 1A). The proportion of incorrect automated rhythm diagnosis was higher in ECGs with poorer signal quality. A regression analysis revealed patient characteristics prone to incorrect heart rhythm diagnosis included male sex (OR 1.61 95% confidence interval (CI) 1.22-2.13, P<0.001), age (per year OR 1.06 95%CI 1.04-1.09, P<0.001), body height (per cm OR 1.02 95%CI 1.00-1.03, P=0.03), and body weight (per kg OR 0.99 95%CI 0.98-0.99, P=0.005). A multivariate model for prediction of correct heart rhythm diagnosis showed a modest area under the curve (AUC) of 0.65 (sensitivity 61%, specificity 58%; Figure 1B). The prediction improved by adding information on signal quality rated by human overread (AUC 0.77, sensitivity 77%, specificity 62%) and even more by the machine-learning based MLSQE score (AUC 0.87, sensitivity 71%, specificity 90%).

Conclusion: The patient characteristics prone to incorrect automated heart rhythm diagnosis of male sex, older age, greater height, and lower body weight probably reflect the target population of many AF screening cohorts. Automated MLSQE scoring for ECG signal quality could improve accuracy in clinical use and potentially reduces the need for cumbersome and time-consuming human overread.
Figure 1  A - Signal quality by human rating and proportion of incorrect automated heart rhythm diagnosis. B - Prediction of correct heart rhythm diagnosis. Blue line - Multivariate model with forward selection considering age, height, weight, body mass index, congestive heart failure, hypertension, diabetes mellitus, vascular disease, stroke/TIA and sex. Purple line - Model with human signal quality (SQ) score. Green line – Model with MLSQE score.