Prediction of incident heart failure through deep learning-based analysis of 1-lead ECGs

M. Neyazi¹, J. Bremer¹, M. Knorr¹, J. Brederecke¹, F. Ojeda¹, S. Blankenberg¹, R.B. Schnabel¹

¹University Medical Center Hamburg Eppendorf, Department of Cardiology, University Heart & Vascular Center Hamburg, Hamburg, Germany

Funding Acknowledgements: Type of funding sources: Public grant(s) – EU funding. Main funding source(s): EU Horizon 2020 AFFECT-EU ERC Consolidator Grant, agreement No 648131

Background/Introduction: Identification of individuals at risk for heart failure (HF) can lead to early intervention. However, large scale screening for incident HF in the general population is insufficient, partly due to the limited capacities in health care systems and number of health care specialists. One-lead ECGs, e.g. from wearables capable of deriving this biomarker, are becoming more popular in the population. This holds the potential to identify individuals at risk who could benefit from a referral to a HF specialist.

Purpose: We examined the ability to predict three-year HF incidence from 1-lead ECGs using a deep learning (DL) model that was trained on data from the UK Biobank (UKB).

Methods: The UKB was established in 2006 and contains data from 500,000 participants from the British general population. N = 100,551 1-lead ECGs from 91,898 unique participants were used for this study. This includes 57,897 3-lead resting ECGs written during ergometry exams between 2006 and 2010 (Exam 1; median age 56.7 years, 46.2% men) and at the first follow-up from 2010-2013 (Exam 2; median age 61.3 years, 49% men) and 32,828 12-lead resting ECGs which were acquired between 2014 and 2020 (Exam 3; median age 63.5 years, 48.4% men). Lead I was extracted from the ECGs for all analyses. A 1D-Convolutional Neural Network was used for prediction of the 3-year risk of incident HF. A comparison with the CHARGE-AF score was performed using a nested 6-fold cross-validation scheme. The models were evaluated using the Area Under the Curve (AUC).

Results: In all three exams, the DL model was able to reliably detect 3-year incident HF (AUC; Exam 1: 0.738, Exam 2: 0.822, Exam 3: 0.738) while the CHARGE-AF score provided slightly better predictions at Exam 1 and Exam 3 (AUC 0.766, 0.748), but lower results in Exam 2 (AUC 0.736). Combining both models led to increased discriminatory properties across all exams (AUC; Exam 1: 0.812, Exam 2: 0.832, Exam 3: 0.788).

Conclusion: The derived DL model shows high discriminatory ability in the prediction of incident HF from 1-lead ECGs. The predictive ability can be further increased when combined with simple clinical variables from the CHARGE-AF score. This study represents another step towards the application of DL-based analysis of the ECG to advance health care.