**Appendix**

**Deep learning-based NT-proBNP prediction from the ECG for risk assessment in the community**

Meraj Neyazi 1,2,3, Jan P. Bremer1,2, Marius S. Knorr1,2, Marcus Vollmer4,5, Stefan Gross5,6, Jan Brederecke1,2, Nils Schweingruber7, Dora Csengeri1,2, Benedikt Schrage1,2, Martin Bahls5,6, Nele Friedrich5,8, Tanja Zeller1,2, Stephan Felix5,6, Stefan Blankenberg1,2, Marcus Dörr5,6, Renate B. Schnabel1,2\*

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

2German Centre for Cardiovascular Research (DZHK), partner site Hamburg/Kiel/Luebeck

3Department of Genetics, Harvard Medical School, Boston, United States

4Institute of Bioinformatics, University Medicine Greifswald

5German Centre for Cardiovascular Research (DZHK), partner site Greifswald

6Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany

7Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

8Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Germany

Meraj Neyazi and Jan P. Bremer contributed equally to this work and share first authorship.

**Address for correspondence:**

Renate B. Schnabel

Postal address: Martinistrasse 52, 20251 Hamburg, Germany

Phone: +49 (0)40 7410 53979

Fax: +49 40 7410 53622

E-Mail: r.schnabel@uke.de

**Supplementary Methods**

*Disease definitions*

The heart failure definition in HCHS is based on the 2021 guidelines of the European Society of Cardiology.(1) For external validation and prospective evaluation of DL-predicted NT-proBNP (pNT-proBNP) for heart failure in the SHIP cohorts, the classification introduced in the Rotterdam Study was used(2). The Rotterdam Criteria have been validated before and classify patients based on their medical history and medication instead of assessment of biomarkers and left-ventricular function. This definition used in the SHIP Study allowed us to utilize cohort data in the absence of consistently performed echocardiography.

AF was defined by medical history in HCHS and defined by an ECG-based diagnosis in SHIP. A history of myocardial infarction (MI) was defined using the medical history in HCHS; in SHIP, ECGs were analysed by trained specialists to identify signs of past MI in addition to self-reported medical history. In SHIP and HCHS, stroke was defined by self-reported medical history. In SHIP, diabetes was defined by self-reported medical history or diabetes medication (ATC code A10) or HbA1c ≥6.5% or serum glucose level ≥11.1 mmol/l at any time, while it was defined in HCHS using the same criteria or a fasting glucose level ≥7.0 mmol/l. In SHIP-START-1 and SHIP-START-2, the participants were not asked for a history of diabetes, but whether diabetes has been diagnosed by a physician since the last SHIP assessment. Mortality information in SHIP was derived from the most recent death certificates. Cardiovascular death was defined in SHIP by ICD10 coding for the main reason of death: I10-I80 or R96.

*Description of deep learning model architecture*

In the proposed architecture, we employ a 1D Residual Neural Network (ResNet1D), implemented using PyTorch and executed on a CUDA-enabled GPU. The model is specifically designed to process unidimensional signals with an input dimensionality of (12,2048), where 12 represents the number of channels and 2048 denotes the number of samples per channel. The network commences with an initial 1D convolutional layer that has a kernel size of 17 and performs downsampling to produce an output with 64 filters and 2048 samples. Following this, the architecture consists of a sequence of five residual blocks. These blocks have varying dimensions, specified as (64,2048), (128,1024), (196,256), (256,64), and (320,16) for the number of filters and samples, respectively. Each residual block encompasses two 1D convolutional layers, complemented by batch normalization and dropout layers with a rate of 0.2.

The architecture concludes with a global average pooling operation along the time dimension, followed by a fully connected (linear) layer that maps the network's output to a single output class. This design choice aligns with the problem's requirement for a single-class output.

The architecture leverages the benefits of residual connections to facilitate the training of deep networks, adhering to the principles laid out in seminal works on deep residual learning.(3,4)

**Supplementary Tables**

**Supplementary Table 1:** Brier-Scores for classification of prevalent heart failure.

|  |  |  |  |
| --- | --- | --- | --- |
| **Cohort** | **pNT-proBNP adjusted for Risk Factors** | **mNT-proBNP adjusted for Risk Factors** | **RF-ModelLogRisk Factor Model** |
|  | **Brier Scores (95% CI)** |
| HCHS | 0.0307(0.0304, 0.0309) | 0.0289(0.0287, 0.0292) | 0.0319(0.0317, 0.0321) |
| SHIP-START-0 | 0.0232(0.0230, 0.0234) | 0.0230(0.0227, 0.0232) | 0.0236(0.0234, 0.0238) |
| SHIP-START -1 | 0.0251(0.0248, 0.0253) | 0.0250(0.0247, 0.0252) | 0.0250(0.0248, 0.0253) |
| SHIP-START -2 | 0.0215(0.0212, 0.0217) | 0.0214(0.0212, 0.0217) | 0.0213(0.0210, 0.0216) |
| SHIP-TREND-0 | 0.0200(0.0198, 0.0201) | 0.0200(0.0198, 0.0202) | 0.0201(0.0200, 0.0202) |

Provided are Brier-Scores for the diagnosis of prevalent heart failure with .95% bootstrapped confidence intervals, supplementary to Table 2. Baseline characteristics in the adjusted models are: age, sex, BMI, hypertension, diabetes, smoking status, LDL-cholesterol, atrial fibrillation, and eGFR.

AUROC = Area Under the Receiver Operating Characteristics Curve, BMI = Body Mass Index, CI = Confidence Interval, eGFR = estimated glomerular filtration rate, LDL = low-density lipoprotein, mNT-proBNP = measured N-terminal prohormone of brain natriuretic peptide, pNT-proBNP = predicted N-terminal prohormone of brain natriuretic peptide.

**Supplementary Table 2**Brier-Scores for incident cardiovascular diseases.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Cohort** | **pNT-proBNP adjusted for Risk Factors** | **mNT-proBNP adjusted for Risk Factors** | **RF-ModelLogRisk Factor Model** |
|  |  | **Brier Scores (95% CI)** |
| Heart Failure | SHIP-START-1 | 0.0162(0.0162, 0.0163) | 0.0161(0.0161, 0.0162) | 0.0162(0.0162, 0.0162) |
| SHIP-START-2 | 0.0365(0.0363, 0.0368) | 0.0356(0.0353, 0.0358) | 0.0373(0.0371, 0.0375) |
| SHIP-START-3 | 0.0627(0.0622, 0.0632) | 0.0613(0.0607, 0.0618) | 0.0641(0.0636, 0.0646) |
| SHIP-TREND-1 | 0.0216(0.0215, 0.0217) | 0.0212(0.0211, 0.0214) | 0.0214(0.0213, 0.0216) |
| Stroke | SHIP-START-1 | 0.0090(0.0090, 0.0090) | 0.0090(0.0090, 0.0090) | 0.0090(0.0090, 0.0090) |
| SHIP-START-2 | 0.0207(0.0206, 0.0208) | 0.0207(0.0206, 0.0208) | 0.0208(0.0207, 0.0209) |
| SHIP-START-3 | 0.0377(0.0374, 0.0380) | 0.0376(0.0373, 0.0380) | 0.0384(0.0381, 0.0387) |
| SHIP-TREND-1 | 0.0179(0.0179, 0.0179) | 0.0179(0.0178, 0.0179) | 0.0179(0.0178, 0.0179) |
| Atrial Fibrillation | SHIP-START-1 | 0.0098(0.0097, 0.0099) | 0.0099(0.0098, 0.0100) | 0.0097(0.0096, 0.0098) |
| SHIP-START-2 | 0.0249(0.0247, 0.0251) | 0.0228(0.0225, 0.0231) | 0.0254(0.0252, 0.0257) |
| SHIP-START-3 | 0.0354(0.0350, 0.0358) | 0.0325(0.0321, 0.0330) | 0.0366(0.0362, 0.0370) |
| SHIP-TREND-1 | 0.0092(0.0091, 0.0092) | 0.0092(0.0091, 0.0092) | 0.0092(0.0091, 0.0092) |
| Myocardial Infarction | SHIP-START-1 | 0.0092(0.0092, 0.0093) | 0.0091(0.0091, 0.0092) | 0.0092(0.0091, 0.0092) |
| SHIP-START-2 | 0.0280(0.0280, 0.0280) | 0.0281(0.0279, 0.0282) | 0.0278(0.0277, 0.0279) |
| SHIP-START-3 | 0.0499(0.0496, 0.0502) | 0.0502(0.0499, 0.0505) | 0.0498(0.0495, 0.0500) |
| SHIP-TREND-1 | 0.0156(0.0155, 0.0157) | 0.0156(0.0155, 0.0157) | 0.0155(0.0155, 0.0156) |
| Cardiovascular Death | SHIP-START-1 | 0.0081(0.0080, 0.0081) | 0.0081(0.0080, 0.0081) | 0.0081(0.0080, 0.0081) |
| SHIP-START-2 | 0.0243(0.0241, 0.0244) | 0.0243(0.0242, 0.0245) | 0.0245(0.0244, 0.0247) |
| SHIP-START-3 | 0.0493(0.0490, 0.0495) | 0.0484(0.0481, 0.0487) | 0.0493(0.0491, 0.0496) |
| SHIP-TREND-1 | 0.0128(0.0127, 0.0128) | 0.0126(0.0126, 0.0127) | 0.0128(0.0127, 0.0128) |

Provided are Brier-Scores for the classification of incident cardiovascular diseases with 95% bootstrapped confidence intervals, supplementary to Table 3. Baseline characteristics in the adjusted models are: age, sex, BMI, hypertension, diabetes, smoking status, LDL-cholesterol, atrial fibrillation, and eGFR. The bootstrapped Brier-Score values of pNT-proBNP adjusted were tested for significance using z-test in comparison to mNT-proBNP adjusted and RF-ModelLogRisk Factor Model.

BMI = Body Mass Index, eGFR = estimated glomerular filtration rate, LDL = low-density lipoprotein, mNT-proBNP = measured N-terminal prohormone of brain natriuretic peptide , pNT-proBNP = predicted N-terminal prohormone of brain natriuretic peptid.

**Supplementary Table 3**Net Reclassification Improvement values for prevalent and incident cardiovascular diseases.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Cohort** | **pNT-proBNP** | **mNT-proBNP** |
| **adjusted for Risk Factors** | **adjusted for Risk Factors** |
|   |   | **Net Reclassification Improvement values (95% CI)** |
| Heart Failure (prevalent) | SHIP-START-0 | 0.0891 | 0.0396 |
| ( -0.1214 - 0.3714) | ( -0.2041 - 0.3162) |
| SHIP-START-1 | 0.08 | 0.0318 |
| ( -0.1272 - 0.3395) | ( -0.2041 - 0.2688) |
| SHIP-START-2 | 0.0082 | 0.0056 |
| ( -0.2023 - 0.2651) | ( -0.1921 - 0.1972) |
| SHIP-TREND-0 | 0.0426 | 0.0359 |
| (-0.1765 - 0.2833) | (-0.1711 - 0.2873) |
| Heart Failure (incident) | SHIP-START-1 | 0.0675 | 0.0374 |
| ( -0.162 - 0.4262) | ( -0.0311 - 0.419) |
| SHIP-START-2 | 0.0646 | 0.045 |
| ( -0.2375 - 0.3846) | ( -0.2308 - 0.3345) |
| SHIP-START-3 | -0.0267 | 0.0196 |
| ( -0.2485 - 0.1248) | ( -0.153 - 0.2008) |
| SHIP-TREND-1 | 0.041 | -0.0045 |
| (-0.2047 - 0.3156) | (-0.2839 - 0.2047) |
| Stroke | SHIP-START-1 | 0.004 | 0.0033 |
| ( -0.0112 - 0.009) | ( -0.0135 - 0.009) |
| SHIP-START-2 | 0.0072 | 0.0387 |
| ( -0.1703 - 0.2857) | ( -0.2703 - 0.4255) |
| SHIP-START-3 | 0.0303 | -0.0018 |
| ( -0.218 - 0.3271) | ( -0.2734 - 0.2268) |
| SHIP-TREND-1 | 0.0234 | 0.0083 |
| (-0.0345 - 0.2547) | (-0.023 - 0.2431) |
| Atrial Fibrillation | SHIP-START-1 | 0.0726 | -0.0091 |
| ( -0.4928 - 0.7451) | ( -0.4952 - 0.25) |
| SHIP-START-2 | 0.0585 | 0.0263 |
| ( -0.34 - 0.4609) | ( -0.2222 - 0.3333) |
| SHIP-START-3 | 0.0588 | 0.018 |
| ( -0.2462 - 0.3903) | ( -0.1818 - 0.2727) |
| SHIP-TREND-1 | 0.1167 | -0.0036 |
| (-0.0181 - 0.7477) | (-0.0113 - 0.009) |
| Myocardial Infarction | SHIP-START-1 | 0.0022 | -0.0031 |
| ( -0.0118 - 0.0118) | ( -0.0118 - 0.0071) |
| SHIP-START-2 | -0.0021 | 0.0087 |
| ( -0.2355 - 0.2355) | ( -0.2289 - 0.2389) |
| SHIP-START-3 | -0.013 | -0.011 |
| ( -0.2264 - 0.1538) | ( -0.2262 - 0.1584) |
| SHIP-TREND-1 | 0.0024 | 0.0137 |
| (-0.1429 - 0.1429) | (-0.1383 - 0.2834) |
| Cardiovascular Death | SHIP-START-1 | 0.0545 | 0.0276 |
| ( -0.4054 - 0.5983) | ( -0.3965 - 0.4054) |
| SHIP-START-2 | -0.0074 | 0.0055 |
| ( -0.2427 - 0.1967) | ( -0.1857 - 0.1948) |
| SHIP-START-3 | 0.0185 | -0.0044 |
| ( -0.109 - 0.1582) | ( -0.1105 - 0.0876) |
| SHIP-TREND-1 | 0.0897 | 0.0043 |
| (-0.2121 - 0.4947) | (-0.2041 - 0.2094) |

Provided are Net Reclassification Improvement values for prevalent and incident cardiovascular diseases with 95% bootstrapped confidence intervals, comparing adjusted models to RF-ModelLog Risk Factor Model. Baseline characteristics in the adjusted models are: age, sex, BMI, hypertension, diabetes, smoking status, LDL-cholesterol, atrial fibrillation, and eGFR.

BMI = Body Mass Index, eGFR = estimated glomerular filtration rate, LDL = low-density lipoprotein, mNT-proBNP = measured N-terminal prohormone of brain natriuretic peptide , pNT-proBNP = predicted N-terminal prohormone of brain natriuretic peptide.

**References**

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42:3599–726.

2. Mosterd A, Hoes AW, de Bruyne MC, Deckers JW, Linker DT, Hofman A, et al. Prevalence of heart failure and left ventricular dysfunction in the general population; The Rotterdam Study. Eur Heart J. 1999;20:447–55.

3. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conf Comput Vis Pattern Recognit CVPR [Internet]. Las Vegas, NV, USA: IEEE; 2016 [cited 2023 Sep 6]. page 770–8. Available from: http://ieeexplore.ieee.org/document/7780459/

4. Kim Y. Convolutional Neural Networks for Sentence Classification. Proc 2014 Conf Empir Methods Nat Lang Process EMNLP [Internet]. Doha, Qatar: Association for Computational Linguistics; 2014 [cited 2023 Sep 6]. page 1746–51. Available from: https://aclanthology.org/D14-1181