

Supplementary Table 2. Summary of performance metrics for all 30 included studies

| Study (Author, Year)            | Primary model (proposed)                     | Comparator models / traditional scores                                                                                   | AUC-ROC                                                  | C-index                                                                         | Accuracy (%)  | Sensitivity (%) | Specificity (%) | Precision    | F1-score     | Recall       |
|---------------------------------|----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|---------------------------------------------------------------------------------|---------------|-----------------|-----------------|--------------|--------------|--------------|
| Agudelo et al., 2025            | Neural Network (best on unbalanced dataset)  | DT, SVM, RF, Gaussian Bayesian Network;<br>Compared against traditional CVD risk scores                                  | 0.690 (95% CI: 0.622–0.759)<br>[unbalanced dataset]      | Not reported                                                                    | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Alaa et al., 2019               | AutoPrognosis (automated ML)                 | Framingham (AUC: 0.724±0.004), Cox PH (0.758±0.005),<br>SVM, RF, NN, AdaBoost, GBM                                       | 0.774 ± 0.005 (95% CI: 0.768-0.780)                      | Not reported                                                                    | <i>NR</i>     | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Amarbayasgalan et al., 2019     | Deep Autoencoder NN (DAE-NNs)                | NB, KNN, DT, RF, SVM, NN;<br>Comparator: Framingham (AUC: 52.2)                                                          | 0.84                                                     | Not reported                                                                    | 83.53%        | Not reported    | Not reported    | 89.56%       | 84.36%       | 79.94        |
| Amarbayasgalan et al., 2019     | Reconstruction Error AE-DNNs                 | PCA-DNNs, NB, KNN, DT, RF, SVM                                                                                           | 0.8665                                                   | Not reported                                                                    | 86.33%        | Not reported    | 90.41%          | 91.37%       | 86.91%       | 82.9         |
| Amarbayasgalan et al., 2021     | Attention-based DNN                          | NB, KNN, DT, SVM, RF, AdaBoost, standard DNN                                                                             | 0.8257                                                   | Not reported                                                                    | 82.78%        | Not reported    | 87.89%          | 86.65%       | 81.66        | 77.27        |
| Ambale-Venkatesh et al., 2017   | Random Survival Forest (RSF) — full model    | AIC-Cox (FW/BW), LASSO-Cox, Cox-PHM;<br>Comparators: ACC/AHA ASCVD (C-index: 0.73),<br>Framingham (C-index: 0.73)        | Not reported                                             | C-index: 0.77–0.86                                                              | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Dehnavi & Shafiee, 2020         | Neuro-Fuzzy model                            | NN variants (Levenberg, Bayesian, Gradient)                                                                              | Not reported                                             | Not reported                                                                    | 65.5–99.5%    | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Dolezalova et al., 2021         | DeepSurv model                               | CPH model, Framingham (C-index: 0.704)                                                                                   | Not reported                                             | 0.745 (95% CI: 0.744–0.745)                                                     | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Goldman et al., 2021            | ANN-CHD (MLP)                                | Framingham Risk Score (FRS-CHD, AUC: 0.837)                                                                              | 0.834 (95% CI: 0.780-0.889)                              | Not reported                                                                    | Not reported  | Not reported    | Not reported    | 0.146–0.247  | 0.2–0.349    | Not reported |
| Hathaway et al., 2021           | DeepSurv                                     | CoxPH, RSF, ISVM, nMTLR                                                                                                  | 0.84 (95% CI: 0.83-0.84) (MAE)                           | Not reported                                                                    | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Hoogveen et al., 2020           | XGBoost — protein model                      | Clinical-risk model (AUC: 0.809±0.057),<br>Combined model (AUC: 0.692±0.090)                                             | 0.705 ± 0.071 (derivation)<br>0.905 ± 0.009 (validation) | Not reported                                                                    | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Johri et al., 2022              | LSTM (with SMOTE)                            | RNN, RF, SVM;<br>Comparators: FRS (AUC: 0.50), ASCVD (AUC: 0.50)                                                         | 0.99 ± 0.01                                              | Not reported                                                                    | 95.80 ± 1.78% | 88.35 ± 5.71%   | 96.66 ± 2.80%   | Not reported | Not reported | Not reported |
| Kasim et al., 2025              | Logistic Regression (with L1 regularisation) | SVM, RF, XGBoost, NB, NN, ensemble models<br>FRS (0.72), RPCE (0.74)                                                     | 0.769 (95% CI: 0.708-0.829)                              | Not reported                                                                    | 73            | 68.3            | 0.732           | 0.085        | 0.151        | 0.683        |
| Li et al., 2024                 | XGBoost                                      | LASSO (C-index: 0.7883),<br>Refitted China-PAR (C-index: 0.7805)                                                         | Not reported                                             | 0.7918 (95% CI: 0.7776-0.8060)                                                  | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Lip et al., 2023                | Neural Network (MLP, 5 hidden layers)        | ML-based Logistic Regression (C-index: 0.674)                                                                            | Not reported                                             | 0.67 (95%CI 0.667–0.674) - training<br>0.668 (95%CI 0.663–0.673) - validation   | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Liu et al., 2025                | MT-BERT (Multitask Transformer)              | CoxPH, RSF, XGBoost, DeepSurv, DeepHit                                                                                   | 0.744 (95% CI: 0.738–0.749)                              | 0.713 (95% CI: 0.710–0.716)                                                     | 82.30%        | Not reported    | 87.40%          | Not reported | Not reported | 0.405        |
| Liu et al., 2025                | Age-specific DeepSurv                        | Traditional DeepSurv (C-index: 0.7358), Simple Cox (C-index: 0.7288)                                                     | Not reported                                             | 0.7332 (95% CI: 0.7253-0.7410) (UKB)<br>0.7409 (95% CI: 0.7359–0.7458) (CHERRY) | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Matheson et al., 2022           | Random Survival Forest (RSF)                 | Indirect comparison to Cox-based models only                                                                             | Not reported                                             | Not reported                                                                    | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Mehrabani-Zeinabad et al., 2023 | Logistic Regression (best performer)         | LDA, ODA, SVM, KNN, DT, RF, ANN, GBM, BART, BARTm<br>variants                                                            | 73.37 ± 0.07                                             | Not reported                                                                    | 64.75 ± 0.23% | 70.23 ± 0.42%   | 63.78 ± 0.30%   | Not reported | Not reported | Not reported |
| Nai-Arun & Mounghai, 2020       | Logistic Regression                          | Back-propagation NN, RF, DT, KNN, NB                                                                                     | Not reported                                             | Not reported                                                                    | 99.94%        | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Neumann et al., 2022            | Cox Proportional Hazards                     | None                                                                                                                     | 68.1 (95% CI: 65.86 -70.36)                              | Not reported                                                                    | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Ordikhani et al., 2022          | XPARS (GA-optimized chart-based model)       | PROCAM (AUC: 0.683), Framingham (AUC: 0.633), PARS<br>(AUC: 0.74), WHO, SCORE;<br>ML: RF (0.712), NB (0.727), CNN (0.74) | 0.739 (XPARS 2D)                                         | Not reported                                                                    | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |
| Shen et al., 2024               | AESurv (deep learning autoencoder survival)  | DeepSurv (C-index SHS: 0.845±0.013), RSF, GBRTs, CPH;<br>External validation: WHI cohort                                 | 0.893 ± 0.009 (SHS)                                      | 0.853 ± 0.010 (SHS)<br>0.752 ± 0.019 (WHI)                                      | Not reported  | Not reported    | Not reported    | Not reported | Not reported | Not reported |

[illegible]