Supplementary Table S3. Data Extraction

Sl. no.	Authors	Objectives	Study Design	Data type	Participants	Outcome	Missing Data	Model Tested	Model Performance	Limitations	Focus Area	Feature extraction and ranking	Cross-validation	Unimodal/Multimodal	Internal/External validation? How and which data?	Prediction (Years)
1	Agudelo et al., (2025)	To develop and evaluate AI-based models for predicting cardiovascular risk in a Colombian population	Retrospective cohort	Patient Data.	874 patients aged 30–74 years who were free of cardiovascular events at baseline and were followed at the Primary Prevention Clinic of the Central Military Hospital, at Bogotá	Cardiovascular events over 10 years	Not mentioned	1.Neural networks 2. Decision trees 3. Support vector machine 4. Random forests 5. Gaussian Bayesian networks	Best: Neural network (AUC ≈ 0.69 unbalanced; 0.67 balanced); others: AUC ~0.56–0.65; Framingham: 0.53	1. External validation was not performed	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	5-cross fold validation	Unimodal	Internal	10 years
2	Ahmed M. Alaa et al. (2019)	**(1)** whether ML techniques based on a state-of-the-art automated ML framework (AutoPrognosis) could improve CVD risk prediction compared to traditional approaches, and **(2)** whether considering non-traditional variables could increase the accuracy of CVD risk predictions.	Prospective Cohort	Registry data. UK Biobank during the period spanning from 2006 to 2010 The minimum follow-up period for all participants was 5 years	Inclusion Criteria - 423,604 participants who were 40 years of age or older and had no known history of CVD (England, Wales, and Scotland) Exclusion Criteria - patients with previous history of coronary heart disease, other heart disease, stroke, transient ischaemic attack, peripheral arterial disease, or cardiovascular surgery	Outcome measure - binary variable, Fatal and non-fatal CVD episode. Clinical Outcome measure - sensitivity and positive predictive value (PPV) Time horizon of 5 years.	Excluded variables that are missing for more than 85 per cent of the population. MissForest - non-parametric data imputation algorithm was used to recover the missing data in the 5 imputed datasets.	(1) ML-based model derived using AutoPrognosis (2) Framingham Risk Score (3) Cox proportional hazards model. (4) Standard ML models - linear support vector machines (SVM), random forest, neural networks, AdaBoost, and gradient boosting machines	AutoPrognonis trained for all the variables (AUC-ROC of 0.774) outperformed the baseline Framingham model and other compared conventional models.	(1) The main limitation of the study is the absence of the cholesterol biomarkers in the UK Biobank data repository, which hindered direct comparisons with the QRISK2 scores currently recommended by the NICE guidelines. (2) Another limitation is that the UK Biobank cohort is ethnically homogeneous: 94% of the participants were of white etnnicity.	Comparison of fine-tuned models with traditional clinical risk scores	Variable Ranking - post-hoc approach. Used RF algorithm for post-hoc variable ranking	10-fold-stratified-cross-validation was implemented to avoid overfitting	Unimodal - most of the variables are non-lab variables collected through questionnaries	Internal	5 years
3	Amarbayasgalan et al. (2019)	Deep autoencoder model of CHD risk prediction - high-risk and low-risk.	Cross-Sectional Survey	Korea National Health and Nutrition Examination Survey (KNHANES) dataset	Included Korean population from KNHSNES dataset. with related features for CHD prediction. 25,990 records including 12,915 records are highly risky people who have the probability to suffer from CHD and 13,075 records are low-risky people	Accuracy, precision, recall, F-measure, and AUC.	Removed rows with missing values.	(1) Data autoencoder based NN (2) Framingham Risk Score (3) Naive Bayes (4) K-nearest neighbors (5) Support Vector Model (6) Decision Tree (7) Random Forest	Experimental results showed that the proposed DAE-NNs outperformed all the compared classifiers with accuracy, precision, F-measure and AUC score of 83.53%, 89.56%, 84.36%, and 84.02%, respectively.	The model cannot handle the missing values	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	10-fold cross-validation	unimodal	Internal	Not mentioned
4	Amarbayasgalan et al. (2019)	Developing a reconstruction error (RE) based deep neural networks (DNNs) to predict high-risk people who are to suffer from CHD.	Cross-Sectional Survey	Korea National Health and Nutrition Examination Survey (KNHANES) dataset The samples spaned from the years 2010-2015	Included Korean population from KNHSNES dataset with related features for CHD prediction. 25,990 records including 12,915 records are highly risky people who have the probability to suffer from CHD and 13,075 records are low-risky people	Accuracy, precision, recall, Specificity, F-measure, and AUC.	Removed rows with missing values.	(1) Autoencoder- DNN (2) PCA-DNN (3) Naive Bayes (4) K-nearest neighbors (5) Support Vector Model (6) Decision Tree (7) Random Forest	Experimental results showed that the proposed AE-DNNs outperformed all the compared classifiers with accuracy, F-measure, and AUC score of 86.33%, 86.91%, and 86.65%, respectively.	Limits to Korean subjects.	Comparison of fine-tuned models with traditional clinical risk scores	Reconstrution errror based feature extraction was done	10-fold cross-validation	unimodal	Internal	Not mentioned
5	Amarbayasgalan et al. (2021)	To develop an attention-based deep neural network (DNN) to improve coronary heart disease (CHD) risk prediction. Compare its performance against traditional machine learning models.	Cross-Sectional Survey	KNHANES dataset	Korean population, classified as low-risk (13,075 records) and high-risk (12,915 records)	Prediction of CHD risk (low-risk vs. high-risk)	Not mentioned	-Attention-based DNN (proposed model) - Traditional models: Naïve Bayes (NB), K-Nearest Neighbors (KNN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), AdaBoost, standard DNN	- Proposed attention-based DNN: Accuracy = 82.78%, F1-score = 81.66%, AUC = 82.57% - Best-performing traditional model (RF): Accuracy = 81.24%, F1-score = 80.02%, AUC not specified	1. Focuses on the Korean population, so generalizability to other populations is uncertain. 2. Uses only selected features from the Framingham Risk Score, potentially missing other relevant predictors.	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	10-fold cross-validation	unimodal	Internal	Not mentioned
6	Ambale-Venkatesh, B., et al. (2017)	To test the ability of random survival forests (RF), a machine learning technique, to predict six cardiovascular outcomes in comparison to standard cardiovascular risk scores.	Prospective, population-based observational cohort	Multi-Ethnic Study of Atherosclerosis (MESA) (USA)	6814 men and women representing four racial/ethnic groups, aged 45–84 years and free of clinical cardiovascular disease	All-cause death, stroke, all cardiovascular disease (CVD), coronary heart disease (CHD), atrial fibrillation (AF) and heart failure (HF) events adjudicated as part of MESA were used as end-points	Not mentioned	1. Random Survival Forest 2. Akaike Information Criterion for Cox regression with backward stepwise elimination (AIC-BW Cox) 3. Akaike Information Criterion for Cox regression with forward stepwise regression (AIC-FW Cox) 4. LASSO-Cox 5. Cox-PHM	C-index	1. No external validation was performed	Direct comparison of existing methods without modification	RF technique was applied to identify the top 20 predictors of each outcome	-	Multimodal	Internal	12-year
7	Dehnavi and Shafiee (2020)	Fuzzy inference to predict the risk of heart disease	Prospective Cohort	Patient Data. Framingham Hospital over 10 years. (Havent mentioned when they started the study) (USA)	500 healthy patients from Framingham Hospital	Predicition of CVD risk	Only patients with complete information are considered. The remaining are dropped.	1. Neuro-Fuzzy 2. Fuzzy interference 3. Neural Network. The Grasshopper Optimisation Algorithm (GOA) was used to optimise the non-linear parameter of the fuzzy set.	Neuro-fuzzy performs better. But the study does not measure any clinical outcomes. The comparison is done only based on the accuracy.	1. Small data size	Development of a new prediction model	Genetic Algorithm was used for feature extraction	-	Unimodal	No validation was performed	Not mentioned
8	Dolezalova N et al. (2021)	To develop a CVD risk model (DiCAVA) applicable in a remote/digital setting using CPH and DeepSurv; to identify novel patient-centric variables for CVD risk assessment; to develop a model viable without cholesterol and blood pressure measurements	Prospective cohort	UK Biobank	466,052 UK Biobank participants (after excluding those with pre-existing CVD); 42,377 (9.09%) developed CVD; median follow-up 11.5 years (IQR 10.7–12.3); age range 37–73 at recruitment (mean 56.2)	Incident CVD	Missing values imputed by mean substitution; variables with <0.2% occurrence excluded; categorical/ordinal variables one-hot encoded; continuous variables scaled	Cox Proportional Hazards (CPH) and DeepSurv (deep neural network Cox model); full model (608 features) and reduced model (47 features after feature selection); additional model excluding cholesterol + SBP (45 features, substituting heart rate for SBP); compared to Framingham risk score and Framingham-retrained CPH	Reduced CPH (47 features): c-index 0.7443 (95% CI 0.7441–0.7445). Reduced DeepSurv: c-index 0.7446 (95% CI 0.7441–0.7452). Model excluding cholesterol/SBP: CPH c-index 0.741, DeepSurv c-index 0.739. AUCC/D ~0.76 from year 6 onwards. Good calibration (ICI: CPH 0.00295, DeepSurv 0.00567). Both superior to Framingham score	No external validation outside UK Biobank; predominantly White UK cohort (limited ethnic diversity); mean imputation for missing data; feature selection performed only on CPH (not DeepSurv due to computational cost); smartphone-accessible use-case excluded some clinical measurements; future studies needed on heterogeneous samples	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	No	Unimodal (structured self-reported and clinical data from UK Biobank)	Internal	10 years
9	Goldman et al. (2021)	To improve coronary heart disease (CHD) risk prediction using an artificial neural network (ANN) model, comparing its performance to the Framingham Risk Score (FRS)	Prospective Cohort	Clinical data from the Framingham Heart Study (FHS) offspring cohort, including health parameters and CHD event records (USA)	Inclusion: 3066 subjects aged 30+ years, from the Framingham Offspring Study Exclusion: Subjects with prior CHD and missing data, patients younger than 30 years	Prediction of CHD within 10 years	47 records with missing data were excluded	ANN with MLP architecture was compared with Framingham risk score	- ANN performed better in lift and gain measures for high-risk percentiles. - ANN had better precision-recall metrics. - FRS had a slightly higher AUC (0.837 vs. 0.834 for ANN). - ANN had better sensitivity and specificity at higher risk levels	1. Results are specific to the Framingham dataset and may not generalize to other populations 2. Need for external validation on diverse datasets	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	5-fold cross-validation	unimodal	Internal	10
10	Hathaway et al. (2021)	CVD risk prediction using a novel deep learning survival model and comparison against existing statistical and machine learning models	Prospective Cohort	Multi-Ethnic Study of Atherosclerosis (MESA) (USA)	Inclusion: Only participants without clinical cardiovascular disease at baseline were included in the study	Prediction of MAE	Median Imputation was performed	1. COXPH 2. RSF 3. Linear SVM (ISVM) 4. Neural multi-task Logistic regression (nMTLR) 5. DeepSurv	- In comparison to the COXPH model, DeepSurv significantly improved ASCVD risk prediction for MAE (AUC: 0.82 vs. 0.80) and mortality (AUC: 0.87 vs. 0.84) with traditional risk factors alone.	1. No external dataset used for validation.	Fine-tuning of pre-trained models with comparison to other ML/DL	Permutation importance and mean decrease Gini	-	Multimodal	Internal and external validation	10
11	Hogeveen et al. (2020)	To hypothesize that a protein-based risk model can outperform prediction using traditional risk factors in the primary prevention setting	Prospective Cohort	Prospective cohort EPIC - Norfolk (nested case-control) and PLIC cohort (validation cohort) (UK and Italy)	822 from the EPIC-Norfolk cohort and 702 from the PLIC cohort.	AUC	Not mentioned	XGBoost A model was trained on the protein data, another on the clinical data, and a third on both protein and clinical data.	The protein model (ROC AUC: 0.905 ± 0.009) was significantly better than the clinical risk model (ROC AUC: 0.609±0.057) in the validation cohort. The combined protein and clinical risk model resulted in an ROC AUC of 0.692±0.090 which was not better than the protein risk model.	1. The cohort study was over a decade old and included individuals with European ancestry. 2. The validation cohort had a limited number of CVD events 3. Targeted proteomics was used instead of untargeted proteomics 4. The model does not predict lifetime risk 5. The samples in the derivation cohort were non-fasting, while the samples in the validation cohort were collected after an overnight fast.	Development of a new prediction model	Not mentioned	5-fold cross-validation	multimodal - proteomics	External validation was performed on the PLIC cohort	<3 years

12	Johri et al. (2022)	To evaluate whether AE3.0DL-based deep learning algorithms, using conventional risk factors and carotid ultrasound imaging, outperform ML-based algorithms and conventional cardiovascular risk calculators (CCVRC) in predicting coronary artery disease (CAD) risk	Cross-Sectional Survey	Retrospective clinical and imaging data, including carotid ultrasound and coronary angiography scores (canada)	500 patients with carotid ultrasound and corresponding coronary angiography scores (CAS), divided into four risk classes (No-Risk, Mild-Risk, Moderate-Risk, High-Risk)	CAD risk classification into four categories based on coronary angiographic scores (CAS)	SMOTE	1. LSTM 2. RNN 3. Random Forest 4. SVM	- LSTM: 95.34% accuracy, AUC = 0.99 - RNN: 95.00% accuracy, AUC = 0.98 - RF: 86.42% accuracy, AUC = 0.91 - SVM: 85.00% accuracy, AUC = 0.90	1. Limited dataset size 2. Need for further validation with cross-modality studies and larger multicenter datasets	Fine-tuning of pre-trained models with comparison to other ML/DL	PCA based polling strategy was used for feature extraction	10-fold cross-validation	multimodal - coronary angiographic score, carotid ultrasound imaging, and clinical data	Internal	Not mentioned
13	Kasim et al., (2025)	To develop and validate ML models for 10-year CVD risk prediction in a Malaysian cohort; to compare performance against FRS and RPCE; to identify key predictors using SHAP analysis; to deploy best-performing model on a web platform	Longitudinal community-based cohort	REDISCOVER Study, Malaysia	5,688 complete-case participants (from 6,130 selected; original cohort 12,617) from 40 communities across Malaysia, aged ≥18 years, no prior CVD at baseline, recruited 2007–2017	Fatal and non-fatal CVD events within 10 years of baseline	442 cases with missing data imputed using MICE (multivariate imputation by chained equations, predictive mean matching); classified as missing completely at random (MCAR)	Base models: 1. SVM (linear & RBF) 2. Logistic Regression (L1 regularization) 3. Random Forest 4. XGBoost 5. Naive Bayes 6. Neural Network. Ensemble/stacked models using GLM, GBM, and RF as meta-learners. Best performer: base LR with LR-selected features	Best model (LR, LR-selected variables): AUC 0.769 (95% CI 0.708–0.829). FRS AUC: 0.716; RPCE AUC: 0.740. NRI improvement: 13.15% vs FRS (p<0.00001), 7.00% vs RPCE. Ensemble models did not outperform best individual model	Data from 2007–2017 (may not reflect current risk factor prevalence); ensemble methods did not outperform simpler LR; focus on 10-year horizon only; no external validation on independent dataset; data not publicly available	Fine-tuning of pre-trained models with comparison to other ML/DL	SHAP analysis identified top predictors	5-fold cross validation	Unimodal (structured clinical/biochemical/demographic data)	Internal	10 years
14	Li et al. (2024)	To leverage electronic health records (EHRs) with irregular longitudinal measurements to improve risk prediction of atherosclerotic cardiovascular disease (ASCVD) using machine learning (ML) models.	Observational Cohort	Chinese Electronic Health Records Research in Yinzhou (CHERRY) study	Inclusion: i) age ranging between 40 and 79 years at the entry time (ii) Chinese residents registered in the health information system during the period between 1 January 2010 and 31 December 2016 (iii) had been living in Yinzhou for at least 6 months. Sample Size - 215,744 Exclusion: i)had no records of serum lipid measurements ii) had a history of CVD before being enrolled in the study	5-year ASCVD risk, defined as non-fatal myocardial infarction, coronary heart disease death, or fatal/non-fatal stroke.	Managed using multiple imputation and machine learning-based missing data handling.	eXtreme Gradient Boosting (XGBoost) and Least Absolute Shrinkage and Selection Operator (LASSO) regression.	XGBoost C-statistic: 0.792 LASSO C-statistic: 0.788	1. Based on the Chinese cohort; need to validate on a different population setting.	Fine-tuning of pre-trained models with comparison to other ML/DL	Not mentioned	5-fold cross-validation	Unimodal	Internal validation	5-year risk
15	Lip et al., (2023)	To examine CVD incidence rates in a Medicare population without prior CVD history; to develop and validate an algorithm predicting incident CVD outcomes using statistical and ML methods; to identify contemporary risk factors in elderly multimorbid population	Retrospective cohort	Medicare health plan, USA, 2016–2022	154,551 individuals; mean age 68.8 years; 62.2% female; 81% aged ≥65 years; no prior CVD for first 3 years; ≥48 months continuous enrollment	Incidence of CVD	Not reported; administrative data used as-is; binary encoding (1=present, 0=absent) for all conditions	Main effect logistic regression (individual risk factors); ML-based logistic regression (main effects + interaction + polynomial/quadratic terms, stepwise selection); Neural network (multilayer perceptron, 5 hidden layers). Decision curve analysis used for clinical utility	Main effect LR: C-index 0.67 (95% CI 0.667–0.674) training; 0.668 (95% CI 0.663–0.673) validation. ML-based LR: C-index ~0.674 training, ~0.68 validation. Neural network: C-index ~0.68	Observational administrative data; potential confounding bias; no blood pressure, lipid, or eGFR values; disease severity unmeasured; residual channeling bias possible; data sharing restricted by corporate agreements	Fine-tuning of pre-trained models without comparison to other ML/DL	Not mentioned	No	Unimodal (structured administrative/claims data)	Internal	Up to ~3.8 years (≥1 year follow-up after 3-year comorbid history; median survival ~21–24 months)
16	Liu et al. (2025)	To develop a hybrid multitask deep learning model (MT-BERT) integrating structured and textual EHR features to predict 10-year CVD risk, enhance individualized stratification, and support equitable assessment across demographic subgroups	Retrospective cohort	Electronic Health Records (EHR) — structured clinical variables and textual representations derived from CPRD Aurum (UK primary care database)	469,496 patients aged 40–85 years from CPRD Aurum (UK); 855,422 valid records after QC; development set n=711,052; spatial external validation set n=144,370 (London practices)	Composite CVD (MI, stroke, TIA, CHD, angina, HF, AAA, PAD), QRISK-defined CVD (CHD, stroke, TIA), plus individual outcomes (CHD, MI, stroke, angina, HF, vascular dementia); binary classification and time-to-event (TTE) over 10 years	Continuous variables imputed using linear regression if ≥3 prior measurements in 2-year span; absence of records interpreted as no occurrence/no event rather than missing	MT-BERT (Multitask Bidirectional Encoder Representations from Transformers): hybrid DistilBERT + MLP with multihead attention and residual connections; custom FocalCoxLoss for joint classification and survival objectives; benchmarked against CoxPH, RSF, GBSA, XGBS, DeepSurv, DeepHit	Composite CVD AUROC: males 0.744 (95% CI 0.738–0.749), females 0.782 (95% CI 0.768–0.796) on test set; C-index: males 0.713, females 0.732; Brier score: males 0.130, females 0.091. Spatial external validation AUROC: males 0.736, females 0.775. MT-BERT outperformed all benchmarks	Computationally intensive; DistilBERT not domain-pretrained (general corpora); C-index consistently lower than AUROC; reduced performance in South Asian and deprived subgroups; no systematic interpretability framework; only spatial external validation, no cross-system validation; fairness mitigation not applied	Fine-tuning of pre-trained models with comparison to other ML/DL	not mentioned	10-fold cross-validation	Multimodal (structured EHR data + textual representations of clinical variables via DistilBERT)	Internal and External validation	10 years
17	Liu et al., (2025)	To propose a population-based recalibration method for survival neural networks and validate it for 10-year cardiovascular risk prediction across different populations (UK and China)	Prospective cohort	UK Biobank and CHERRY	UK Biobank: 347,206 individuals (training + internal validation); CHERRY cohort: 177,756 individuals (external validation)	First-ever cardiovascular disease event	participants with missing predictors excluded	DeepSurv, age-specific DeepSurv, DeepHit (SNNs); Cox model; Cox with age interactions	Discrimination: C-index >0.72 across models; best ~0.7409 (age-specific DeepSurv); calibration improved after recalibration; initial underprediction ~60% in external cohort	1. Used a completecase analysis, which may introduce selection bias.	Fine-tuning of pre-trained models without comparison to other ML/DL	Not mentioned	No	Unimodal	Internal and External validation	10
18	Matheson, M. B., et al. (2022)	To identify important markers of incident CVD in the low-risk populations and to generate an algorithm through which individualized risk predictions for up to 6 years may be calculated.	Retrospective Cohort	Structured data including biomarkers, health history, medication use, and lifestyle variables. (Japan)	155108 adults aged 40 years and more were included.	Prediction of CAD and ASCVD risk	Dynamic Imputation	Random Survival Forest	AUC: 84% (CAD) and 82% (ASCVD)	1. No external validation 2. Population specific	Fine-tuning of pre-trained models without comparison to other ML/DL	Variable importance via Minimum depth of maximal subtree	-	Unimodal	Internal	1,3,5
19	Mehrabani-Zeinabad et al., (2023)	To develop classification models using ML and statistical techniques for predicting future CVD incidence; to identify the most efficient predictors of CVD; to compare ML with traditional statistical methods in an eastern Mediterranean population	Longitudinal community-based cohort	Isfahan Cohort Study, Iran, 2001–2017; 16-year follow-up	5,432 participants (from 6,323 recruited) aged ≥35 years, free of CVD at baseline, Isfahan, Iran	Any CVD event during 16-year follow-up: fatal/non-fatal MI, fatal/non-fatal stroke, sudden cardiac death, unstable angina. CVD events occurred in 819 participants (15.08%)	Variables with >10% missing excluded (385 retained for most models); MissForest (non-parametric RF-based) used to impute remaining missing values for most models. BARTm handles up to 90% missing natively without imputation	LR, LDA, QDA, SVM, kNN, Decision Tree, Random Forest, ANN, GBM, BART, BARTm.10% (up to 10% missing), BARTm.90% (up to 90% missing). Grid search cross-validation for hyperparameter tuning	QDA: highest accuracy 75.50%; DT: highest sensitivity 82.52%. LR and ANN: highest AUROC (~73.37% and ~73.35%), highest balanced accuracy (~67.00%). BARTm.90%: accuracy 69.48%, sensitivity 54.00%. MCC: ANN and LR best (~63%). DT: weakest overall	Loss to follow-up; absence of HbA1c and blood type variables; high missing values in some variables; only 8 variables selected may underfit complex CVD risk; Iranian-specific results limit generalizability: smoking not selected as predictor (low female smoking rate in region)	Fine-tuning of pre-trained models with comparison to other ML/DL	Recursive Feature Elimination (RFE) selected 8 optimal variables	10-fold cross validation	Unimodal (structured clinical/demographic/lifestyle data)	Internal	Not mentioned
20	Nai-Arun and Moungmai (2020)	To predict the risk of CVD	Cross-Sectional Survey	Hospital Data. From the saraburi province Thailand.	31929 participants. Inclusion: Age 35 or more Exclusion - incomplete data	Outcome - CVD risk to five levels based on the Thai Ministry of Public Health. Clinical outcome - MSE and ROC.	The participants with no data were dropped	1) Logistic Regression 2) Back propogation NN 3) K-nearest neighbors 4)Naive-bayes 5) Decision Tree 6) Random Forest	Best performance was by Logistic regression and lowest was Naive-Bayes.	1) The accuracy rate of the logistic regression is very high, which can be because in the preprocessing they removed the incomplete information participants 2) Data collected only from Saraburi, limiting generalisability	Fine-tuning of pre-trained models with comparison to other ML/DL	Pearson Correlation was used to select features based on their relation	10-fold cross-validation	unimodal	Internal	Not mentioned
21	Neumann et al. (2022)	To develop a risk prediction model for incident MACE from subjects enrolled in a large clinical trial in initially healthy, elderly individuals and to validate the model in a large primary care dataset.	Prospective Cohort - PREDICT Randomized Controlled Trial - ASPREE	Aspirin in Reducing Events in the Elderly (ASPREE) study - Derivation Cohort PREDICT cohort - External Validation cohort USA and NewZealand	19,114 community-dwelling individuals aged ≥ 70 years (≥ 65 for US minorities), without prior CVD events, dementia, or physical disability were recruited in Australia and the USA.	Incidence of CVD over the period of 5 years	The IRSAD score was missing in all US participants and was therefore only considered in sensitivity analyses.	Cox Proportional Hazard Regression Model	The model had an AUC of 67.5 in internal and 64.2 in external validation.	1. the number of events in the derivation dataset was lower than would be expected in a general population. 2. analyses were derived from a mainly Caucasian population and thus care may be needed in applying the model to other populations.	Fine-tuning of pre-trained models without comparison to other ML/DL	LASSO	-	unimodal	Internal and External validation	5 year
22	Ordikhani et al., (2022)	To develop a novel explainable CVD risk score (XPARS) using a Genetic Algorithm (GA) optimized chart-based model; to balance interpretability and accuracy; to compare against conventional chart-based (WHO, PARS, SCORE) and ML models	Longitudinal community-based cohort	Isfahan Cohort Study, Iran, 2001–2017; 16-year follow-up	5,432 non-CVD participants aged 35–84 at baseline; 705 CVD cases during 10-year follow-up; 51.3% female, 48.7% male	Fatal and non-fatal CVD events over 10 years: sudden cardiac death, unstable angina, MI, stroke	Not explicitly stated; Cox regression used for initial chromosome population; imbalanced dataset (705 CVD vs 4,727 non-CVD) handled implicitly through GA fitness function (AUROC)	XPARS (GA-optimized chart-based model): 2D variant (8 features, AUROC 0.76), 1D variant (4 features, AUROC 0.72). Compared against: FRAMINGHAM (AUROC 0.633), PROCAM (AUROC 0.683), PARS (AUROC 0.74), WHO, SCORE; ML benchmark: 3-layer CNN (AUROC 0.74)	XPARS 8-feature (2D): AUROC 0.76 (training 0.80, cross-validated 0.76). XPARS 4-feature (1D): AUROC 0.72. Both outperform Framingham, PROCAM, PARS, CNN, and WHO models	Small sample size (5,432; only 107/160 chart blocks had data); Iranian population-specific, limits generalizability; one block trained at a time (may miss cross-block interactions); no external validation; imbalanced dataset not explicitly corrected	Comparison of fine-tuned models with traditional clinical risk scores	Information Gain and Gain ratio for feature significance	10-fold cross validation	Unimodal (structured clinical/demographic data)	Internal	10 years
23	Shen et al. (2024)	To develop a deep learning autoencoder survival analysis model (AESurv) for early prediction of coronary heart disease (CHD) using DNA methylation and clinical data.	Prospective Cohort - SHS Observational Cohort - WHI	1. Strong Heart Study (SHS) 2. Women's Health Initiative (WHI) (USA)	Inclusion: Subjects with no coronary heart disease or missing data for risk factors of cardiovascular disease at baseline, but with available blood DNA methylation measures. 2321 - Strong Heart Study 2107 - Women's Health Initiative	Time-to-event CHD prediction, including fatal and non-fatal myocardial infarction (MI) and CHD-related death.	Few missing CpG sites in the WHI dataset were imputed with K-nearest neighbors algorithm	Deep learning AESurv model utilizing a supervised autoencoder combined with the average negative log partial likelihood loss function from CPH.	AESurv (Best Model) in SHS: C-index = 0.864 ± 0.009, AUROC = 0.905 ± 0.009.	Two distinct populations were used	Development of a new prediction model	Not mentioned	5-fold cross-validation	Multimodal data - used clinical data and DNA methylation	External validation using Women Health Initiave.	Not mentioned

24	Steinfeldt et al. (2022)	Development and validation of a novel neural network-based cardiovascular disease risk model, NeuralCVD	Prospective Cohort	UK Biobank	Inclusion: a cohort of 273,383 women and 229,122 men aged between 37 and 73 years at their baseline assessment. Exclusion: Individuals in the UK Biobank population who withdrew consent, with missing information about their sex or with earlier records of incident myocardial infarction or stroke or lipid-lowering treatment at baseline were excluded. Patients with pre existing myocardial infarction, stroke, or lipid lowering therapy were excluded from the analysis	10 year cardiovascular disease risk defined by the earliest recorded event of fatal or nonfatal myocardial infarction	Multiple Imputations using Chained equations with RF was used to handle missing data	NeuralCVD	- NeuralCVD score outperformed SCORE with a difference in C index of 0·037, ASCVD with 0·024, and QRISK3 with 0·010	1. The UK Biobank study cohort is of generally lower risk for cardiovascular events than the general primary care population and recalibration with a relevant data source should be performed before public application. 2. Although discrimination, reclassification, and calibration are crucial criteria for evaluating predictive models and allow comparison over the established baselines, they are not quantifying an absolute clinical impact at the population level	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	22-fold nested cross validation	unimodal	Internal	10 year
25	Teshale et al. (2024)	To demonstrate the advantages of employing artificial intelligence (AI) models to improve the prediction of CVD risk	Randomized Controlled Trial	Aspirin in Reducing Events in the Elderly (ASPREE) study (USA)	Relatively healthy adults aged 70 years or more (8,122 men, 10,426 women)	Prediction of 10-year atherosclerotic cardiovascular disease (ASCVD) risk, including coronary heart disease-related death, non-fatal myocardial infarction, and ischemic stroke.	Not mentioned	1. DeepSurv 2. Neural MultiTask Logistic Regression 3. Random Survival Forest 4. Cox Proportional Hazards	DeepSurv & NMTLR: Best predictive performance	1. The population was predominantly White and socioeconomically advantaged, limiting generalizability. 2. No external validation was performed.	Fine-tuning of pre-trained models with comparison to other ML/DL	Permutation-based feature importance	5-fold cross-validation	Unimodal	Internal	3-year, 5-year and 10-year
26	Weng et al. (2024)	To develop a deep learning model (DLS) using photoplethysmography (PPG) to predict the 10-year risk of major adverse cardiovascular events (MACE) without requiring physical or lab measurements	Prospective Cohort	UK Biobank	The participants were recruited to the UKB study between 2006 and 2010 Inclusion: For each participant, only measurements related to their first visit were included. Exclusion: Participants with non-fatal myocardial infarction or stroke before their first visit, or missing any of the variables for the model. Participants without body mass index (BMI) or systolic BP (SBP) for a fair comparison against the other office- and lab-based risk prediction models	Prediction of 10-year MACE risk (myocardial infarction, stroke, cardiovascular death).	Patients with missing values were excluded from the study	1. A one-dimensional ResNet18 model was trained for feature extraction to learn PPG representations. 2. A Cox proportional hazards regression model was developed to predict the 10-year MACE risk. For comparisons, different survival models were developed based on different feature sets	1) Non-inferior to the WHO score 2) The sensitivity and specificity are non-inferior	1) Specific population (Non-generalisable) 2) PPG was collected from a specific clinical device; smartphone-based PPG validity needs further study.	Development of a new prediction model	Deep learning-based feature extractor using one-dimensional ResNet18.	-	Unimodal	External	10-year
27	Weng, S. F., et al. (2017)	The aim of this study was to evaluate whether machine-learning can improve accuracy of cardiovascular risk prediction within a large general primary care population.	Prospective Cohort	Routine Clinical Data (UK)	378,256 patients from UK family practices, free from cardiovascular disease at outset.	The primary outcome was the first recorded diagnosis of a fatal or non-fatal cardiovascular event documented in the patient's primary or secondary care computerised record.	Median Imputation	1. Logistic Regression 2. Random Forest 3. Neural Networks 4. Gradient Boosting Machines	Neural networks performed the best, with predictive accuracy improving by 3.6%.	1. No external validation was performed	Fine-tuning of pre-trained models without modifications with comparison to other ML/DL	Not mentioned	2-fold cross validation	Unimodal	Internal	10-year
28	You et al. (2023)	ML-based algorithm for CVD risk prediction that outperformed previously established scores.	Prospective Cohort	Longitudinal cohort study. UK Biobank	Inclusion: 473 611 CVD-free participants aged between 37 and 73 years old. Exclusion: Participants without follow- up records and with pre-existing MI or stroke	10-year incident CVD risk prediction, including myocardial infarction (MI), ischemic stroke (IS), and hemorrhagic stroke (HS).	Variables with more than 70% missing data were excluded	eXtreme Gradient Boosting Machine, random forest, logistic regression, K-nearest neighbors, support vector machine, artificial neural networks Light Gradient Boosting Machine - was called UKCRP, was the best performing one In addition, the SHAP algorithm was implemented to interpret the relationship between the predictors and model output	UKCRP observed an AUC of 0.762±0.010 on 10-year incident CVD, which is significantly greater than those of existing risk scales of QRISK V.3, SCORE V.2, ASCVD, and FGCRS	1. UK Biobank individuals suffer a lower CVD risk relative to the general primary care population 2. Despite that the UKCRP model was well calibrated over spatially different recruitment centres, its value in the pragmatic clinical application should be verified in entirely independent prospective cohorts to ensure that such implementation does improve patient outcomes 3. Because the population of the UK Biobank is predominantly white, the generalisability of the model across ancestrally distinct individuals will help to determine whether more appropriate and ethnically relevant decisions are required.	Comparison of fine-tuned models with traditional clinical risk scores	Not mentioned	20-fold cross-validation	Unimodal	Internal	10-year risk
29	Yu et al. (2024)	To develop a deep learning model (Dynamic-DeepHit) to predict 10-year atherosclerotic cardiovascular disease (ASCVD) risk by incorporating longitudinal risk factor data. Compare its performance with the Pooled Cohort Equations (PCE)	Prospective Cohort	Longitudinal cohort (USA) Cardiovascular Lifetime Risk Pooling Project (LRPP): the Framingham Heart Study, Framingham Offspring Study, Coronary Artery Risk Development in Young Adults (CARDIA) Study, and Atherosclerosis Risk in Communities (ARIC) Study	15565 participants were over 40 and under the age of 75 years with no record of self-report or diagnosed ASCVD at the index exam or during the 8 year observation period, and had at least one measurement of SBP, DBP, total cholesterol and HDL cholesterol.	The outcome in the study was ASCVD incidence over a 10-year period that began at the end of the observation period	Not mentioned	Dynamic-DeepHit model - consists of 2 neural networks 1. RNN 2. fully-connected NN The model also uses an attention mechanism that identifies important longitudinal measurements when making risk predictions	Dynamic-DeepHit AUROC: 0.815 (95% CI: 0.782–0.844).	1. the cohorts used in this study may not reflect the clinical conditions of present-day patients, who are more likely to be on CVD treatments, such as statins. 2. data was recorded more sparsely in the cohort studies	Fine-tuning of pre-trained models with comparison to other ML/DL and traditional risk scores	Not mentioned	Not mentioned	Unimodal	Internal validation - testing data which was not part of model developmnent is used for validation./	10-year follow-up period
30	Zhao, J., et al. (2019)	To demonstarte the performance of machine learning and deep learning on longitudinal EHR data for the prediction of 10-year CVD event, compared to a gold standard achieved by (ACC/AHA) Pooled Cohort Risk Equations	Longitudinal Data in EHR	De-identified copy of the EHRs at the Vanderbilt University Medical Center, USA	Inclusion: Subjects between the age of 18 to 78. 109,490 individuals, including 9,824 cases and 99, 666 controls were a part of the cohort. Exclusion: Individuals with any CVD diagnosis prior to baseline date were excluded; for all individuals, the inpatient physical and laboratory measures were excluded.	Predicition of a 10-year CVD event.	Median Imputation	Machine Learning 1. Logistic Regression 2. Random Forest 3. Gradient Boosting Trees Deep Learning: 1. Convolution Neural Network 2. Long-short Term Memory Network	1. The ACC/AHA equation achieved an average AUROC of 0.732 2. Machine learning models using only ACC/AHA features obtained an AUROC of 0.738–0.751 3. Machine learning models using aggregate EHR features achieved an AUROC of 0.765–0.782. 4. By using temporal features, LR, GBT and deep learning models improved the AUROC to 0.781–0.790.	1. The study was restricted to data obtained during routine clinical practice. 2. Only 204 SNPs were used in the genetic experiment 3. Some of the effects of the SNPs may also be modeled directly by phenotypes.	Direct comparison of existing methods without modification with other ML/DL models and traditional risk scores	Chi-square was used for feature selection methods to select independent features on EHR data	10-fold cross validation for model selection 5-fold cross validation for top ten feature selection	Unimodal	Internal	10-year