# *Integrating genetics, metabolites, and clinical characteristics in predicting cardiometabolic health outcomes using machine learning algorithms - a systematic review*

Computers in Biology and Medicine 186 (2025) 109661

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

# Integrating genetics, metabolites, and clinical characteristics in predicting cardiometabolic health outcomes using machine learning algorithms – A systematic review

Xianyu Zhu [a], Eduard F. Ventura [b], Sakshi Bansal [a], Anisha Wijeyesekera [c], Karani S. Vimaleswaran [a,d,*]

[a] Hugh Sinclair Unit of Human Nutrition, Department of Food and Nutritional Sciences and Institute for Cardiovascular and Metabolic Research (ICMR), University of Reading, Reading, RG6 6DZ, UK
[b] Institute of Agrochemistry and Food Technology-National Research Council (IATA-CSIC), Department of Biotechnology, Av. Agustin Escardino 7, 46980, Valencia, Spain
[c] Food Microbial Sciences Unit, Department of Food and Nutritional Sciences, School of Chemistry, Food and Pharmacy, University of Reading, Reading, RG6 6DZ, UK
[d] Institute for Food, Nutrition and Health (IFNH), University of Reading, Reading, RG6 6AH, UK

ARTICLE INFO

*Keywords:*
Cardiometabolic health
machine learning
Type 2 diabetes
Blood pressure
genetics
Metabolomics

ABSTRACT

*Background:* Machine learning (ML) integration of clinical, metabolite, and genetic data reveals variable results in predicting cardiometabolic health (CMH) outcomes. Therefore, we aim to (1) evaluate whether a multi-modal approach incorporating all three data types using ML algorithms can improve CMH outcome prediction compared to single-modal or paired-modal models, and (2) compare the methodologies used in existing prediction models.

*Methods:* We systematically searched five databases from 1998 to 2024 for ML predictive modelling studies using the multi-modal approach for CMH outcomes. Risk-of-bias assessment tools were used to assess methodological quality. Study characteristics, ML algorithms, data preprocessing, evaluation methods and metrics, feature selections, and feature importance parameters were synthesized narratively to show methodological heterogeneity.

*Results:* Of the four included studies (3 ML algorithms), three were at low risk of bias, and one was at high risk. The multi-modal approach consistently improved T2D and BP prediction compared to single-modal or paired-modal models. Genetics showed the lowest predictive performance in three studies. Logistic regression (n = 2 studies) and random forest (n = 1) were used in T2D studies, while XGBoost was used in one BP study. One study with missing data and variations in feature selection across all studies hindered a comprehensive comparison of feature importance.

*Conclusions:* Our review emphasizes the potential improvement in T2D and BP prediction using ML algorithms with the multi-modal approach. However, further studies using diverse ML algorithms with optimized methodologies on single-modal, paired-modal, and multi-modal models are needed to gain insights into biomarker selection for predicting CMH outcomes.

## 1. Introduction

Cardiometabolic health (CMH) refers to the overall welling of cardiovascular system (i.e., heart and blood vessels) and metabolic system (i.e., how body processes nutrients and energy) [1–3]. CMH comprises a group of health outcomes, including type 2 diabetes (T2D), obesity, insulin resistance (IR), elevated blood pressure (BP), and dyslipidemia [4–8]. With the rising global prevalence of these outcomes [4–6], it is critical to unravel the pathophysiological mechanisms underlying CMH. Predicting CMH outcomes relies on various data, including clinical characteristics, genetic variants, and metabolite concentrations. Clinical characteristics are considered the standard for predicting T2D [9,10], cardiovascular disease (CVD) [11], and BP [12]. Meanwhile, CMH outcomes have also been predicted using metabolites identified with

* Corresponding author. Hugh Sinclair Unit of Human Nutrition, Department of Food and Nutritional Sciences and Institute for Cardiovascular and Metabolic Research (ICMR), University of Reading, Reading, RG6 6DZ, UK.
*E-mail address:* v.karani@reading.ac.uk (K.S. Vimaleswaran).

https://doi.org/10.1016/j.compbiomed.2025.109661
Received 14 April 2024; Received in revised form 2 January 2025; Accepted 6 January 2025
Available online 11 January 2025

metabolomics approaches [13–17], as well as single nucleotide polymorphisms (SNPs), the most common genetic variants used in the calculation of genetic risk scores (GRS) or polygenic risk scores (PRS), identified through genome-wide association studies (GWAS) [18–22]. Nevertheless, some studies have shown that genetics [23,24] may offer only modest improvements in predictive performance. On the other hand, combining clinical characteristics with genetics [18,25,26] or with metabolites [27] has shown improved accuracy in predicting T2D [18,26] or CVD [25,27]. Nonetheless, some studies have suggested that combining clinical characteristics with genetics may not improve the predictive performance of CVD [18,26] or T2D [28]. This observation may be due to heterogeneity in ethnicities [29,30], prediction models [31], and feature selection [32].

Machine learning (ML) has become essential in predictive analytics due to its robust handling of high-dimensional omics data and ability to capture complex non-linear patterns [33]. In contrast, traditional statistical methods focus on inferring relationships between variables but often require limiting variables to avoid overfitting, constraining comprehensive association measurements [33]. In clinical settings, two types of ML algorithms, including supervised learning (i.e., random forest (RF), support vector machines (SVM), multi-layer perceptron (MLP), Bayesian models, regressions, gradient boosting, and neural networks) and unsupervised learning (i.e., principal component analysis (PCA), k-means, and hierarchical clustering), are widely used for predicting CVD [34–36], T2D [37,38], BP [39], and other clinical outcomes [40]. Various evaluation methods, including train-test split [34,38], cross-validation [35,37], and independent cohorts [35,38], are used to mitigate the risk of overfitting and assess the performance of prediction models. Hyperparameter tuning is further complemented with the evaluation methods to optimize predictive performance [34,35,38]. Additionally, the evaluation metrics for classification (i.e., area under the curve (AUC) [34,35,37–40], sensitivity [34,35,38–40], specificity [34,35,38,40], precision [39,40], accuracy [39,40], and F1 score [39,40]) and regression (i.e., $R^2$ [36], root mean square error (RMSE) [39], and mean absolute error (MAE) [39]) are used to quantify the performance and effectiveness of prediction models.

Combining clinical characteristics and metabolites, or using them separately, has shown promising predictive performance, while using genetics alone or in combination with clinical characteristics has shown inconsistent results. This systematic review aims to (1) evaluate whether a multi-modal approach—combining clinical characteristics, genetics, and metabolites—using ML algorithms can improve CMH outcome prediction compared to single-modal or paired-modal approaches, and (2) compare the methodologies used in existing prediction models. To our knowledge, this is the first systematic review to address these specific questions.

## 2. Methods

### 2.1. Search strategy

This systematic review follows the PRISMA guidelines [41] (Table A1). The study protocol was registered on PROSPERO (https://www.crd.york.ac.uk/PROSPERO/), with registration number CRD42023465645. Reviewers (X.Z., E.V.F., A.W., and K.S.V.) collectively developed the search strings, which can be found in Table A2. Next, one reviewer (X.Z.) performed a comprehensive literature search across multiple databases, including SCOPUS, PubMed, Web of Science, Science Direct, and Semantic Scholar, from January 1998 to September 2023. Furthermore, new publications alerted from each database were reviewed weekly until September 25, 2024, to expand the search.

### 2.2. Inclusion and exclusion criteria

Criteria for inclusion and exclusion of published literature are presented using the PICO framework [42] (Table A3). Briefly, this review included observational and interventional studies conducted in human populations and written in English. Studies that used a multi-modal approach—combining clinical characteristic, genetic, and metabolite data—to compare with single-modal or paired-modal approaches using ML algorithms for predicting cardiovascular and metabolic-related outcomes were included (Fig. 1). Definitions for single-modal, paired-modal, and multi-modal approaches described in this systematic review are presented in Table 1. Studies that did not identify genetic variants and metabolites using genetics and metabolomics approaches, respectively, were excluded. Moreover, since the concept of the metabolome was introduced in 1998, studies published before then were excluded [43].

### 2.3. Study selection and data extraction

The search results, including citation details, abstracts, and keywords, were imported into Zotero Software (Version 6.0.26) to facilitate the manual removal of duplicates. Subsequently, the results were imported into Rayyan software [44] for screening and extraction. Two reviewers (X.Z., S.B.) independently screened study titles and abstracts against eligibility criteria prior to full-text screening, and data from eligible studies was extracted. Appendix documents and cited references from selected studies [45–48] were thoroughly reviewed to ensure data consistency and accuracy. Any discrepancies between reviewers were resolved by a third reviewer (K.S.V.).

All available studies that integrated clinical characteristics, genetics, and metabolites using ML algorithms to predict CMH outcomes were extracted according to the CHARMS checklist [49]. While clarifications were sought from the corresponding authors for two studies [45,48], some missing data remained unobtainable from the authors [46,47]. In one study [47], metabolites identified by RF or logistic regression (LR) were compared for "predictive ability" (AUC: 0.923 vs. 0.908, respectively). Nevertheless, since the study did not present metabolite data in their figure, "predictive ability" could not be interpreted in the context of T2D prediction with metabolite data alone. Given the provided metabolite risk scores (available in their appendix files), based on five metabolites identified by RF, we conducted a receiver operating characteristic (ROC) analysis in IBM SPSS Statistics (Version 27) to predict T2D incidence, yielding an AUC of 0.923 (Fig. A1). Consequently, it is reasonable to interpret "predictive ability" as referring to T2D prediction with metabolite data alone.

### 2.4. Assessment of risk of bias

To evaluate the methodological quality and risk of bias (RoB) in the selected articles, we used the Appraisal tool for Cross-Sectional Studies (AXIS) [50] for cross-sectional studies (Table A4), the Newcastle-Ottawa Scale (NOS) [51,52] for case-control studies (Table A5), Non-randomized Studies – of Interventions (ROBINS-I) [53] for cohort studies (Table A6), and the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [54] for ML prediction models (Table A7).

### 2.5. Data synthesis and reporting

The TRIPOD statement [55] was used for reporting adherence (Table A8). A narrative synthesis compiled information from eligible studies [45–48] on study design, outcomes, population, sample size, sex ratio, age, clinical characteristics, genetic variants, metabolites, and prediction scores (Table 2). There was inconsistency in definiting clinical parameters across the literature (e.g., clinical characteristics [45,46,48], clinical factors [46], demographic factors [46], traditional risk factors [48], and biochemical measures [48]). For simplicity and clarity, we used "clinical characteristics" to collectively represent these variables in this review, with original definitions listed in brackets under the column "Clinical characteristics" in Table 2. We also synthesized various feature importance metrics, including odds ratio (OR), 95 % confidence

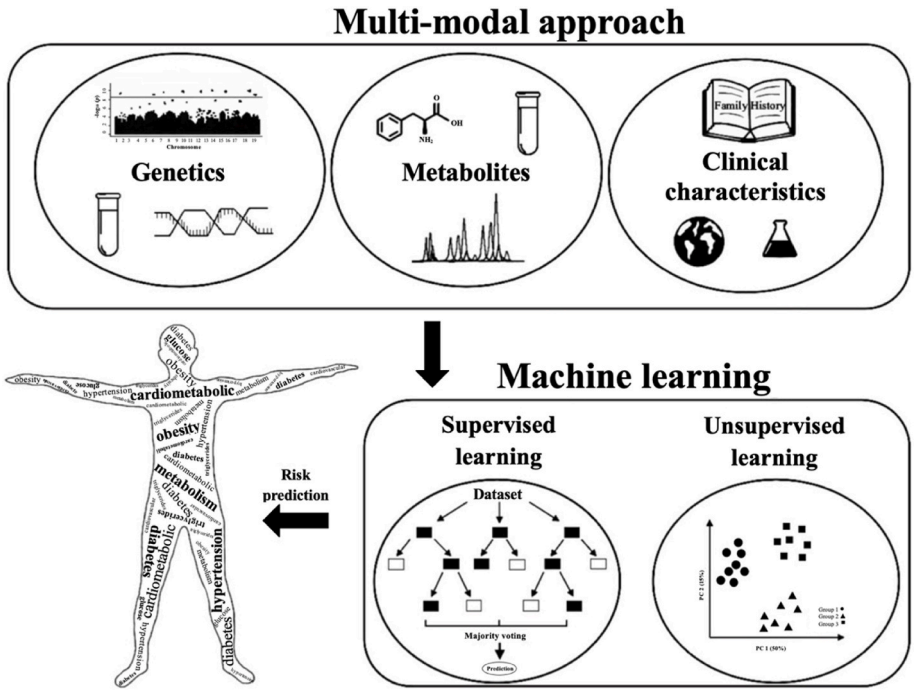


**Fig. 1.** A schematic representation of the study selection for this systematic review. This review included studies that used a multi-modal approach—combining genetics, metabolites, and clinical characteristics—to compare with single-modal (i.e., genetics, metabolites, or clinical characteristics) or paired-modal approaches (i.e., combinations of two data types) using machine learning algorithms to predict cardiometabolic health outcomes.

**Table 1**
Definitions for single-modal, paired-modal, and multi-modal approaches.

| References | Single-modal approach | Paired-modal approach | Multi-modal approach |
|---|---|---|---|
| [45] | Clinical characteristics, genetics, or metabolites | Clinical characteristics + genetics; Clinical characteristics + metabolites; Genetics + metabolites | Genetics + metabolites + clinical characteristics |
| [46] | Clinical characteristics (clinical factors + demographic factors) | Clinical characteristics + genetics | Genetics + metabolites + clinical characteristics |
| [47] | Clinical characteristics, genetics, or metabolites | – | Genetics + metabolites + clinical characteristics |
| [48] | Clinical characteristics (traditional risk factors + biochemical measures) | Clinical characteristics + metabolites | Genetics + metabolites + clinical characteristics + diet |

Note: the definition for each model also reflects the models used in each study.

interval (CI), *p*-values, mean decrease impurity (MDI), and feature orders based on Shapley additive explanations (SHAP) values (Table 3). Prediction models, including their evaluation methods and metrics, are summarized in Table 4. Data preprocessing and feature selection methods for clinical characteristics, genetics, and metabolites are described in Table A9. Reported genetic variants from each study are summarized in Table A10.

## 3. Results

### 3.1. Study selection and risk of bias

The PRISMA flow diagram illustrating the study selection is presented in Fig. 2. A total of 7728 articles were identified using the search strings provided in Table A1. After a thorough review, nine papers were selected for full-text screening. However, five of these were excluded for specific reasons: one study [56] did not include metabolites and genetics in the prediction model; one [57] focused on clustering T2D patients by disease progression rather than using a multi-modal approach to predict T2D outcomes; one [58] did not provide results, and two [59,60] integrated other data (i.e., transcriptomics, methylation, or proteomics) in the prediction model. Consequently, adhering to our predefined exclusion and inclusion criteria, four studies [45–48] with three ML models (i.e., LR [45,46], RF [46], and XGBoost [48]) were selected for this systematic review. The selected studies included two prospective cohort studies [45,46], one nested case-control study [47], and one cross-sectional study [48]. Three studies [45,47,48] were overall at low RoB (Table A4–7), and one study [46] was at high RoB for the prediction model (Table A7). All studies had low applicability concerns (Table A7).

### 3.2. General characteristics of the studies

Table 2 presents an overview of the main characteristics of the included studies. All studies [45–48] contained sample sizes that were above the recommended threshold for prediction models [49]. Sex ratios were balanced in the T2D studies [45–47] and in the Qatari cohort within the UK study [48]. Notably, in the TwinsUK cohort of the BP study, 92.8 % of the participants were females [48]. All studies included age groups from 40 to 67 [45–48]. Evaluation metrics such as AUC [45–47] and $R^2$ [48] were reported for the multi-modal approach. Notably, two studies did not report metrics for the single-modal models (i.e., genetics or metabolites [46,48]; Table 1). Nevertheless, metrics for paired-modal models (i.e., genetics + metabolites [45], clinical + genetics [45,46], and clinical + metabolites [45,48]) were reported

**Table 2**
Characteristics of the studies included in the systematic review.

| Study | Outcome | Population & Study design | Sample size & Sex ratio & Age (±SD) | Clinical characteristics* | | Genetic variants | | Metabolites | | All |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | AUC/ $\underline{R^2}$ | GRS/PRS | AUC/ $\underline{R^2}$ | | AUC/ $R^2$ | AUC/$\underline{R^2}$ |
| [45] | T2D | The US Prospective cohort – 13.5-year follow-up | n = 1622 (206 (12.7 %) new case) 48.1 % Male vs 51.9 % Female 54.7 ± 9.63 | A = (Clinical factors) Age; BMI; Sex; Family T2D history; SBP; FPG; TAG; HDL-C | A: LR 0.856 | B = (GRS) 62-SNP GRS | (B): LR 0.641 (A + B): LR 0.861 | C = (Metabolites) Tyrosine; Isoleucine; Phenylalanine; LPC a C18:2; C38:6 PC; C44:1 TAG; C48:0 TAG; C52:1 TAG; C56:9 TAG | (C): LR 0.803 (B + C): LR 0.820 (A + C): LR 0.874 | (A – C): LR 0.880 |
| [46] | T2D | Korea Prospective cohort – 10-year follow-up | n = 1425 (331 (23.2 %) new case) 45.2 % Male vs 54.8 % Female 55.9 ± 8.8 | A = (Demographic factors) Age; BMI; Sex; Family T2D history; HTN; Smoking; Alcohol B = (Clinical factors) HbA1C; FPG; TAG; HDL-C; Total cholesterol | (A): RF 0.613 LR 0.608 (A + B): RF 0.844 LR 0.835 | C = (PRS) 239,062-SNP PRS | – | D = (Metabolites) Spermine; Isoleucine; Hexose; Valine; Glycine; Alanine; Leucine; LPC a C18:2; PC ae C42:4; PC ae C42:0; PC ae C36:3; PC ae C34:3; PC ae C42:1; PC ae C40:5; PC ae C44:6 | – | (A – C): RF 0.876 LR 0.871 (A – D): RF 0.883 LR 0.875 |
| [47] | T2D | China Nested case-control – 12-year follow-up | n = 220/220 (220 (50.0 %) new case) 41.8 % Male vs 58.2 % Female 53.35 ± 6.71 | A = (Clinical characteristics) BMI; DBP; SBP; FPG; TAG; HDL-C; Waist circumference | (A): LR 0.798 | B = (GRS) 20-SNP GRS | (B): LR 0.586 | C = (Metabolites) Riboflavin; Cnidioside A; 2-methoxy-5-(1H-1, 2, 4-triazol-5-yl)- 4-(trifluoromethyl) Pyridine; 7-methylxanthine; Mestranol | (C): RF 0.923 LR 0.908 MRS* 0.923 | (A – C): LR 0.960 |
| [48]* | BP | The UK Cross-sectional | n = 4863 7.2 % Male vs 92.8 % Female 53.46 ± 13.2 | A = (Traditional risk factors) Age; BMI; Sex B = (Biochemical measures) Urate; Glucose; Phosphate; Chloride; Creatinine; Calcium; Potassium | (A): $\underline{33.0\ \%}$ (A + B): $\underline{37.8\ \%}$ | C = (PRS) 891-SNP PRS | – | D = (Metabolites) 206 metabolites | (A + B + D): $\underline{38.8\ \%}$ | E = (Diet) (A – E): $\underline{39.2\ \%}$ (UK) (A + B + D): $\underline{45.2\ \%}$ (Qatar – external validation) |

[48]* = different hyperparameters applied to corresponded models except traditional risk factors and the multi-modal approach; Clinical characteristics* = original definition of clinical characteristics for each study was included in the brackets; MRS* = extracted from the original paper [47] to confirm whether the authors used RF and LR models to predict T2D incidence with metabolites alone. Abbreviation: AUC = area under the curve; BP = blood pressure; BMI = body mass index; DBP = diastolic blood pressure; FPG = fasting plasma glucose; GRS = genetic risk scores; HTN = hypertension; HDL-C = high-density lipoprotein cholesterol; LR = logistic regression; LPC = lysophosphatidylcholine; MRS = metabolites risk score; PC = phosphatidylcholine; PRS = polygenic risk scores; RF = random forest; SD = standard deviation; SPB = systolic blood pressure; SNP = single nucleotide polymorphism; TAG = triacylglycerol; T2D = type 2 diabetes.

(Table 1). Furthermore, the feature importance of the integrated multi-modal approach was either calculated [45,46] or interpreted [48] in three studies, while one study [47] did not report on this aspect (Table 3, Table 4). In terms of prediction models, data preprocessing, evaluation methods, and metrics, most studies [45–48] reported these aspects, but one study [47] did not specify how data was preprocessed (Table A9) and which prediction model was used (Table 4).

### 3.3. T2D studies: comparison of predictive performances and feature importance

In terms of predictive performances, three studies [45–47] conducted in the US, Korea, and China consistently showed that the multi-modal approach improved T2D incident prediction compared to single-modal or paired-modal models (Table 2). When predicting T2D incidence using single models, two studies [45,47] showed that genetic data had the lowest predictive performance compared to clinical or metabolite data (Table 2). Notably, the US study [45] showed higher predictive performance for clinical data over metabolite data, while the Chinese study [47] showed the opposite trend (Table 2). On the other hand, the Korean study [46] added each component incrementally (i.e., clinical characteristics, genetics, and metabolites; Table 2), making it challenging to compare the contribution of single-modal models. Nonetheless, we can deduce that clinical characteristics have higher predictive performance than genetics or metabolites, given their highest feature importance in the multi-modal approach [46] (Table 3).

In comparing the relative importance of variables in the multi-modal approach across two T2D studies [45,46], fasting plasma glucose (FPG) consistently emerged as the most predictive clinical characteristic (Table 3). Notably, no SNPs or metabolites showed consistent patterns across these studies [45,46] (Table 3).

**Table 3**
Summary of the important features determined within the context of integrated multi-modal approach in each study and shared features across the studies.

| | | | (Clinical characteristics + Genetic variants + Metabolites) | | | | | |
|---|---|---|---|---|---|---|---|---|
| Study | Cohort | Outcome | Clinical characteristics | | Genetic variants | | Metabolites | |
| | | | OR (95 % CI; *P*)/MDI/ Feature order* (No._) | Shared | OR (95 % CI; *P*)/ MDI | Shared | OR (95 % CI; *P*)/MDI/Feature order* (No._) | Shared |
| [45] | The US | T2D | FPG<br>1.10 (1.08, 1.13; <0.0001)<br>Family T2D history<br>2.07 (1.36, 3.16; 0.001)<br>HDL-C<br>0.98 (0.96, 0.99; 0.01) | FPG | 62-SNP GRS<br>1.06 (1.02, 1.10; 0.001) | – | Phenylalanine<br>3.12 (1.36, 7.13; 0.007)<br>C52:1 TAG*<br>1.90 (1.19, 3.04; 0.01)<br>LPC a C18:2*<br>0.32 (0.14, 0.75; 0.01)<br>C38:6 PC*<br>0.23 (0.07, 0.71; 0.01)<br>C44:1 TAG*<br>0.23 (0.07, 0.71; 0.01) | – |
| [46] | Korea | T2D | HbA1C<br>1.88 (1.61, 2.19; <0.0001)/0.1869<br>FPG<br>1.55 (1.27, 1.89; <0.0001)/0.1398 | | 239,062-SNP PRS<br>2.05 (1.77, 2.37; <0.0001)/ 0.1545 | | Spermine<br>0.82 (0.71, 0.96; 0.010)/0.0509<br>Isoleucine<br>1.32 (1.02, 1.86; 0.026)/0.0299 | |
| [48]* | The UK | BP | (Traditional risk factors)<br>1_Age; 2_BMI<br>(Biochemical Measures)<br>4_Urate; 7_Glucose; 9_Phosphate; 10_Chloride | Age; BMI Urate | – | – | 3_Dihomo-linolenate; 5_Cis-4-decenoyl carnitine; 6_Lactate; 8_Cortisol | Dihomo-linolenate; Cis-4-decenoyl carnitine; Lactate; Cortisol |
| [48]* | Qatar | BP | (Traditional risk factors)<br>1_Age; 2_BMI; 3_Sex<br>(Biochemical measures)<br>5_Urate | | – | | 4_Dihomo-linolenate; 6_Cis-4-decenoyl carnitine; 7_Lactate; 8_Phenylacetylglutamine; 9_Cortisol; 10_Histidine | |

Note: The Chinese study [47] was not included due to missing data [48];* = only top 10 features were compared; Feature order* = orders are corresponding to the SHAP values; Abbreviation: BP = blood pressure; BMI = body mass index; CI = confidence interval; FPG = fasting plasma glucose; GRS = genetic risk scores; HbA1c = hemoglobin A1c; HDL-C = high-density lipoprotein cholesterol; MDI = mean decrease impurity; OR = odds ratio; PC = phosphatidylcholine; PRS = polygenetic risk scores; LPC = lysophosphatidylcholine; SNP = single nucleotide polymorphism; TAG = triacylglycerol; T2D = type 2 diabetes.

### *3.4. BP study with two cohorts: comparison of predictive performances and feature importance*

Unlike T2D studies [45–47], which assessed important features after building all prediction models, one BP study [48] first identified important features (n = 50 features) from a multi-modal approach (n = 264 variables), integrating clinical characteristics, dietary intake, genetic, and metabolite data in the TwinsUK cohort [48]. Subsequently, given that genetic data failed to feature in the top 50 variables, genetic data was excluded during external validation in the Qatari cohort [48] and during incremental testing in the TwinsUK cohort [48]. Dietary intake data was also excluded in the Qatari cohort [48].

As for predictive performances, the TwinsUK cohort [48] showed improved model performance in predicting BP with the multi-modal approach, surpassing the performance of both single-modal and paired-modal models (Table 2). In addition, external validation in the Qatari cohort [48] showed even higher model performance (Table 2). However, it is noteworthy that the TwinsUK cohort [48] was characterized by middle-aged females (i.e., 53.46 ± 13.2; Table 2), whereas the Qatari cohort [48] had a relatively younger demographic (i.e., 39.11 ± 12.0 [48]) with a balanced sex ratio (i.e., 50.2 % male, 49.8 % female; Table 2). Although the study [48] did not conduct single-modal models to predict BP in the TwinsUK cohort, the incremental testing showed that clinical characteristics had the highest predictive performance, whereas adding metabolites to clinical characteristic data resulted in only 1 % improvement (Table 2).

The top 50 influential features were identified in both the UK and Qatari cohorts [48]. Seven of the top 10 features, including three clinical characteristics (i.e., age, body mass index (BMI), and urate) and four metabolites (i.e., dihomo-linolenate, cis-4-decenoly carnitine, lactate, and cortisol), overlapped between these two cohorts (Table 3). While age and BMI played prominent roles in shaping the prediction model in both cohorts, it is noteworthy that sex had a higher impact on predictions in the Qatari cohort compared to the TwinsUK cohort. This discrepancy may be attributed to differences in the sex ratio between the two cohorts [48].

### *3.5. Prediction models and their evaluation methods and metrics*

While LR [45,46] was the prevailing model for T2D prediction, the XGBoost [48] was used for BP prediction (Table 4). Notably, the Korean study [46] used four ML algorithms (i.e., LR, RF, XGBoost, and MLP) and established that the RF-based model provided the most accurate prediction for T2D (Table 4). Furthermore, hyperparameter tuning was not performed [45] or reported [47] in two studies (Table 4).

K-fold cross-validation was the primary evaluation method in three studies [46–48], while the Jackknife technique was used in one study [45]. Additionally, two studies [46,47] combined k-fold cross-validation with bootstrap, albeit with variations in their specific methodologies (Table 4). Nevertheless, only two studies [45,48] specified a train-test split ratio, and one [48] validated prediction performance in a separate cohort (Table 4). Moreover, the interpretation of feature importance in the UK study [48] was gauged through SHAP values and visualized by a PCA biplot [48].

For evaluation metrics, AUC held prominence in all T2D studies [45–47], with the net reclassification improvement (NRI) used in two studies [45,46] (Table 4). On the other hand, the UK study used $R^2$ to measure the explained variance of BP. Feature importance was

**Table 4**
Summary of the methodologies for the machine learning prediction models of cardiometabolic health outcomes from the studies.

| Study | ML-based prediction model | | | Evaluation method | | Evaluation metrics | | | | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| | Algorithm | Hyperparameter tuning | Feature importance | Internal validation | External validation | Discrimination/ General metrics | Reclassification metrics | Others | Feature importance | 95 % CI measurement |
| [45] | LR | – | Calculated | The Jackknife technique with 10 random samples of 90 % of cohort; Split ratio (90:10) | – | AUC (entire cohort) | NRI | – | OR | Jackknife resampling |
| [46] | LR; RF | Tree-structured Parzen estimator-based Bayesian optimization for LR; 10-fold stratified cross-validation for RF | Calculated | 10-fold nested cross-validation with 100 bootstrap replicates; Split ratio (not reported) | – | AUC (average); Brier score; Log-loss | NRI; cNRI; IDI | ROC curve; Precision-recall curve; NBD curve; Calibration curve | OR for LR; MDI for RF | Not clear |
| [47] | Not reported | Not reported | Not reported | 10-fold cross-validation with 2000 bootstrap replicates; Split ratio (not reported) | – | AUC (average); Specificity; Sensitivity; Youden index | – | – | Not reported | (Bootstrapping) 2000 bootstrap replicates of the 10-fold cross-validation |
| [48] | XGBoost | Grid search | Interpreted with SHAP values and PCA biplot | 5-fold cross-validation in TwinsUK; Split ratio (80:20) | Replicated in QBB | $R^2$; MAE; MAPE | – | Calibration curve (scatter plot) | – | Standard deviation of the 5-fold cross-validation |

Note: although both studies [46,47] combined k-fold cross-validation with bootstrap, bootstrapping was performed within each fold for AUC calculation [46] or used after k-fold cross-validation to compare models' AUCs [47]. Abbreviation: AUC = area under the curve; cNRI = category-free NRI; CI = confidence interval; IDI = integrated discrimination improvement; LR = logistic regression; ML = machine learning; MAE = mean absolute error; MAPE = mean absolute percentage error; MDI = mean decreased impurity; NBD = net benefit-based decision; NRI = net reclassification improvement; OR = odds ratio; PCA = principle component analysis; QBB = Qatari biobank; RF = random forest; ROC = receiver operating characteristic; SHAP = Shapley additive explanation.

calculated using OR for LR [45,46] and MDI for RF [46] (Table 4). Diverse techniques, including the Jackknife [45], the bootstrap [47], and the standard deviation of the 5-fold cross-validation [48], were used to estimate the 95 % CI for AUC [45,47] or $R^2$ [48]. Nonetheless, the method for estimating the 95 % CI was not specified in the Korean study [46] (Table 4). Only two studies [46,48] performed calibration curve to assess the reliability of the prediction model.

### 3.6. Overview of data preprocessing and feature selections for clinical characteristics, genetics, and metabolites

Two studies [45,46] included data without missing values. One study [48] used various imputation methods, including KNN, twin genotype imputation, minimum imputation, and exclusion of 20 % missing values (Table A9). One study [47] did not report data preprocessing. The selection of clinical characteristics varied across studies, with some studies relying on previous association studies [45,46,48] or baseline *p*-values [47] (Table A9). Moreover, the selection of SNPs was based on previous GWAS related to T2D [45,47] or BP [48], with one study [46] conducting an ethnic-specific GWAS analysis (Table A9). The effect sizes of GRS or PRS on CMH outcomes were reported using β-coefficient [45,48] or OR [46], but it was not reported for one study [47]. Furthermore, the analysis of metabolites used diverse separation techniques combined with mass spectrometry (Table A9). The selection of metabolites also involved various approaches, including LR-based backward stepwise [45], the Boruta algorithm along with a sensitivity analysis [46], a combined LR and RF technique [47], and the inclusion of known metabolites [48] (Table A8).

## 4. Discussion

This systematic review highlights the advantages of integrating genetics, metabolites, and clinical characteristics as a multi-modal approach using ML algorithms to improve CMH prediction, outperforming both single-modal and paired-modal models. This integrated approach using ML applications could provide insights into better biomarker selection, thus improving CMH predictions in clinical practice. However, the high cost of genetic and metabolite tests in research and clinical practice needs to be addressed with more advanced and efficient technologies to promote precision medicine and nutrition. Furthermore, the limited number of CMH studies and varied methodologies used across them [45–48] present challenges in defining the optimal approach for predicting CMH outcomes. Thus, several solutions and suggestions are discussed in this systematic review.

The reviewed studies encompass both categorical (i.e., T2D [45–47]) and continuous (i.e., BP [48]) outcomes within diverse populations, including the US [45], Korea [46], China [47], and the UK [48]. External validation was also conducted on the Qatari population [48]. In both

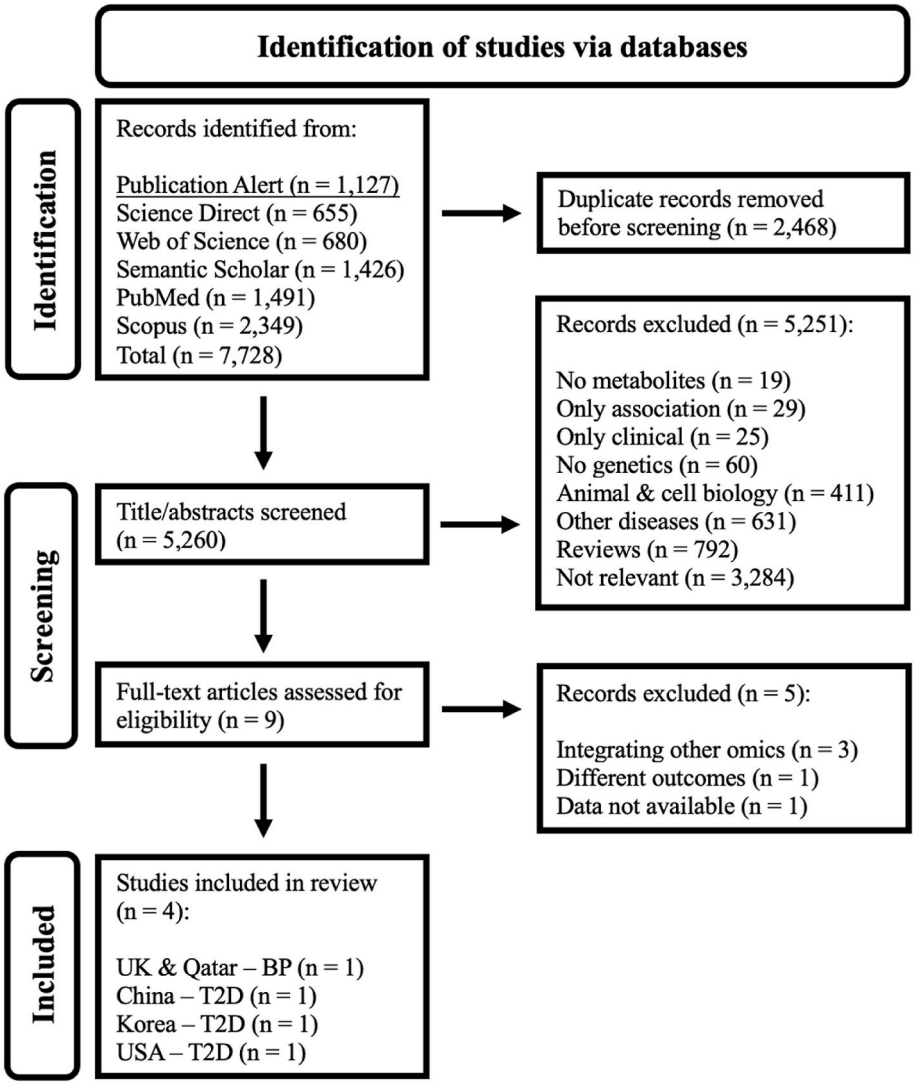


**Fig. 2.** PRISMA flow chart showing the selection of articles for this systematic review. The literature search was conducted in five databases, including SCOPUS, PubMed, Web of Science, Science Direct, and Semantic Scholar, from January 1998 to September 2024. Four papers were included in this systematic review after identification and screening. Reasons for excluded papers were listed. T2D = type 2 diabetes; BP = blood pressure.

cases, the results consistently showed that the multi-modal approach using ML algorithms improved predictive performance compared to single-modal [45–48] or paired-modal models [45,46,48]. This finding aligns with recent studies using a similar multi-modal approach with various ML algorithms to predict other disease outcomes, including rheumatoid arthritis (RA) [61], polycystic ovary syndrome (PCOS) [62], non-alcoholic fatty liver disease (NAFLD) [63]. These studies [61–63] highlight the importance of adopting a multi-modal approach using ML algorithms in predicting disease incidence.

Regarding model contribution to the multi-modal approach, our findings [45,46,48] suggest that clinical characteristics contribute the most to the multi-modal approach, even though one study [47] may have a differing view, possibly due to variations in ethnicities and metabolites [64,65]. On the other hand, genetics [45,47,48] showed the lowest contribution, despite one study [46] holding a different perspective. It is noteworthy that PRS consists of numerous SNPs, while GRS comprises hundreds or fewer [66]. Therefore, the PRS data used in the Korean study [46], with over 200,000 SNPs, is expected to have higher predictive performance than the GRS-based data in the US [45] and Chinese [47] studies, which include fewer than 100 SNPs. This is because the PRS data incorporates most, if not all, heritable variants. However, a recent study [60] showed that while the PRS, consisting of 721,911 variants, improved T2D prediction in individuals with normoglycemia, those with a high PRS in the normoglycemic group still had a low absolute risk of developing T2D. Hence, caution is advised in interpreting the results with high predictive performance based on numerous SNPs. Notably, in the cases of RA, PCOS, and NAFLD, GRS-based data showed higher predictive performance than metabolite data when combined with clinical data [61–63], highlighting the varying predictive performance across different disease outcomes.

Among the influential variables in the T2D prediction models [45,46], FPG emerged as an important determinant within clinical characteristic data, which is unsurprising, given its fundamental role as a diagnostic criterion for T2D [67]. In addition, the absence of consistent SNPs or metabolites across the T2D studies [45,46] may be attributed to genetic heterogeneity [29,30], varied feature selection methodologies [32], diverse ethnic backgrounds and metabolite profiles [64,65], as well as different metabolomics approaches (i.e., targeted vs. untargeted; Table A9) [68]. Given that GRS or PRS primarily correlate with beta-cell function [45,46,69–71], and that metabolites are strongly associated with insulin resistance [57,58,69,72–74], the inclusion of both genetic and metabolite profiles remains essential in predicting T2D incidence. For BP prediction [48], clinical characteristics, particularly age and BMI, emerge as important determinants, aligning with existing literature

[75–78]. Genetic data exert a limited impact on the multi-modal approach for BP prediction [48], which may be attributed to a small fraction (less than 5 %) of the explained variability in hypertension [79,80]. Moreover, metabolite data only modestly improves the predictive performance of the multi-modal approach by 1 % [48]. In contrast, literature [59] showed comparable predictive performance between a single-modal model (metabolite) and paired-modal models (clinical + PRS), suggesting the importance of including metabolite data in predicting BP [31,81,82]. Overall, additional studies investigating BP through a multi-modal approach with ML algorithms are required to validate whether including genetic and metabolite profiles can improve predictive performance.

Various ML algorithms (i.e., LR [45,46], RF [46], and XGBoost [48]) were used to predict CMH outcomes, with LR being the primary model. Given the intricate pathophysiological mechanisms underpinning CMH outcomes, using LR along with other algorithms (e.g., RF, decision trees, neural networks, and SVM) may better capture both linear and non-linear patterns [83–86]. For data preprocessing, deletion [46,48] is a common imputation method for missing data under 5 % [54,87]. However, future research should use multiple imputation for clinical data to achieve the least biased result [54]. Other imputation methods should also be considered for genetic [88] and metabolite data [89] to improve predictive accuracy. For feature selection, previously associated biomarkers [45,46,48] used for clinical characteristics may suffice, but including biochemical tests using ML techniques would add values and prevent overfitting. GRS calculated with previously identified SNPs [45,47,48] may suffice, as PRS does not necessarily outperform GRS in predicting CMH outcomes [45–48,60]. Given the dynamic nature of metabolites, which can change rapidly in response to environment and genetics, multivariate tests (i.e., Boruta [46], RF [47]) may better capture complex relationships [90]. Additionally, least absolute shrinkage selection operator (LASSO) could be performed before RF for initial selection to prevent overfitting [90]. For model performance evaluation, three studies [46–48] used cross-validation methods, with two [46,48] incorporating hyperparameter tuning to optimize predictive performance and mitigate overfitting [34,35,91]. On the other hand, the Jackknife technique was performed in one study [45], which typically does not align with hyperparameter tuning, as its primary purpose is resampling [92]. While the Jackknife technique is more suited for assessing bias and variance in small datasets [92], the standard k-fold cross-validation is preferred for larger datasets [93,94]. Nevertheless, the combination of cross-validation with bootstrap [46,47], particularly nested cross-validation [46], may be favored as it enhances model stability and robustness of confidence intervals [95–97], providing a more comprehensive assessment of model performance. Regarding evaluation metrics, all studies used appropriate metrics, including AUC [45–47] and $R^2$ [48]. While AUC measures binary outcome discrimination, multiple metrics (i.e., specificity, recall, accuracy, precision, F1 score) offer a more comprehensive performance interpretation [98]. For continuous outcomes, while $R^2$ explains the variance captured by different models, other metrics such as MAE and RMSE should also be reported to measure prediction accuracy [39]. Notably, only two studies [46,48] used a calibration curve to assess the reliability of the prediction model. Future research should include calibration curve, as poorly calibrated prediction model may lead to misleading clinical decisions [99]. As for feature importance, OR was used for LR [45,46], MDI for RF [46], and SHAP for XGBoost [48]. OR, commonly used in LR, provides insights into the individual effects of predictors on odds, although it is not directly comparable across predictors or models [100]. On the other hand, both MDI and SHAP offer valuable insights into feature importance in prediction models. MDI, commonly used in ensemble models, provides the collective impact of features by aggregating impurity reduction across all individual models [101]. SHAP is suitable for multiple models and assesses individual contributions by calculating the average impact of each feature across all model combinations [102].

Among the reviewed studies, three [45,46,48] had a low RoB, while one [47] had a high RoB due to missing data. All studies had low applicability concerns. However, the reviewed studies have several limitations. First, the absence of profiles for data preprocessing, feature importance, prediction models, and SNP weighting in one study [47] posed a challenge to a more comprehensive study comparison. Second, studies should also consider transparency when reporting 95 % CI measurements [46] and train-test split ratios [46,47] to enhance reproducibility. Lastly, while all studies reported prediction scores for the multi-modal approach, some did not assess genetic or metabolite data alone as single-modal models [46,48], nor did they examine other combinations of two data as paired-modal models [46–48]. This omission hinders insights into the complementarity and synergy of combining various data.

Our systematic review comes with several limitations. First, this systematic review does not encompass other omics approaches, such as transcriptomics and proteomics, which could offer additional biological insights into CMH complexity. However, our review prioritized model simplicity for broader applicability across diverse settings rather than detailed biological deciphering. Second, given the intricate pathophysiological mechanisms associated with CMH, we exclusively focused on outcomes that describe cardiovascular-metabolic interactions. Accordingly, outcomes such as PCOS [103], NAFLD [104], and chronic inflammatory diseases [105], despite their components impacting both cardiovascular and metabolic health, were not within the scope of this review. Third, we did not assess decision curve analysis for the RoB assessment, as it was not a main component in TRIPOD and PROBAST guidelines. However, future research should include this analysis to optimize clinical decision-making [106]. Fourth, rapidly evolving ML applications may introduce selection bias, meaning the newest ML applications might have been overlooked during screening. However, our focus on multi-modal or omics approaches makes this unlikely. By prioritizing these integrated approaches, we have minimized the risk of selection bias. Lastly, limited literature on CMH-related studies led to a focus on T2D predictions [45–47], constraining our insights into BP or other CMH prediction. However, this review adheres to PRISMA and TRIPOD guidelines, ensuring a comprehensive and comparative analysis that addresses critical knowledge gaps within the field.

## 5. Conclusions

Despite one study with a high RoB, the overall evidence suggests that a multi-modal approach—combining clinical, metabolite, and genetic data—using ML algorithms can improve predictive performance over single-modal or paired-modal approaches. While we have reviewed potential solutions to the challenges in CMH prediction, future research should remain cautious of these limitations and open to more advanced methodologies. Notably, adherence to the TRIPOD guideline is crucial when developing multi-modal CMH prediction models. Additionally, employing various ML applications to identify the best algorithm for CMH prediction, along with reporting feature importance and prediction scores for single-modal, paired-modal, and multi-modal models, can optimize biomarker selection. Such a strategy holds the potential to enhance model robustness, facilitate more accurate predictions, and deepen our understanding of CMH's biological foundations, thereby advancing precision medicine and precision nutrition.

## CRediT authorship contribution statement

**Xianyu Zhu:** Writing – review & editing, Writing – original draft, Validation, Methodology. **Eduard F. Ventura:** Writing – review & editing, Methodology. **Sakshi Bansal:** Writing – review & editing, Validation. **Anisha Wijeyesekera:** Writing – review & editing, Supervision, Methodology. **Karani S. Vimaleswaran:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

## Ethics approval

Not applicable.

## Ethics statement

Not applicable (this is a systematic review)

## Funding

None.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We sincerely thank authors in the reviewed papers, including Jose C. Florez, Bianca C. Porneala, Panayiotis Louca, and Cristina Menni, for providing missing data and clarifications. E.F.V thanks to a Predoctoral grant awarded by the Spanish Ministry of Science and Innovation and the European Social Fund Plus (ESF+) for the training of doctors within the framework of the State Plan for Scientific, Technical and Innovation Research 2021–2023. (ref. CEX2021-001189-S-20-1). E.F.V., IATA-CSIC author, also acknowledges the Spanish government MCIN/AEI to the Center of Excellence Accreditation Severo Ochoa (CEX2021-001189-S/MCIN/AEI/10.13039/501100011033).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.109661.

## References

[1] F. Guo, D.R. Moellering, W.T. Garvey, The progression of cardiometabolic disease: validation of a new cardiometabolic disease staging system applicable to obesity, Obesity 22 (2014) 110–118, https://doi.org/10.1002/oby.20585.
[2] N.E. Boutagy, A.K. Singh, W.C. Sessa, Targeting the vasculature in cardiometabolic disease, J. Clin. Invest. 132 (2022), https://doi.org/10.1172/JCI148556.
[3] J.R. Petrie, T.J. Guzik, R.M. Touyz, Diabetes, hypertension, and cardiovascular disease: clinical insights and vascular mechanisms, Can. J. Cardiol. 34 (2018) 575–584, https://doi.org/10.1016/j.cjca.2017.12.005.
[4] M. Vaduganathan, G.A. Mensah, J.V. Turco, V. Fuster, G.A. Roth, The global burden of cardiovascular diseases and risk, J. Am. Coll. Cardiol. 80 (25) (2022) 2361–2371, https://doi.org/10.1016/j.jacc.2022.11.005.
[5] N.W.S. Chew, C.H. Ng, D.J. Tan, G. Kong, C. Lin, Y.H. Chin, et al., The global burden of metabolic disease: data from 2000 to 2019, Cell Metabol. 35 (3) (2023 Mar), https://doi.org/10.1016/j.cmet.2023.02.003.
[6] M. O'Hearn, J. Liu, F. Cudhea, R. Micha, D. Mozaffarian, Coronavirus disease 2019 hospitalizations attributable to cardiometabolic conditions in the United States: a comparative risk assessment analysis, J. Am. Heart Assoc. 10 (2021) e019259, https://doi.org/10.1161/JAHA.120.019259.
[7] Y. Yoshida-Montezuma, E. Stone, S. Iftikhar, V. De Rubeis, A.T. Andreacchi, C. Keown-Stoneman, et al., The association between late preterm birth and cardiometabolic conditions across the life course: a systematic review and meta-analysis, Paediatr. Perinat. Epidemiol. 36 (2022) 264–275, https://doi.org/10.1111/ppe.12831.
[8] N.E. Zois, E.D. Bartels, I. Hunter, B.S. Kousholt, L.H. Olsen, J.P. Goetze, Natriuretic peptides in cardiometabolic regulation and disease, Nat. Rev. Cardiol. 11 (2014) 403–412, https://doi.org/10.1038/nrcardio.2014.64.
[9] A. Leong, N. Daya, B. Porneala, J.J. Devlin, D. Shiffman, M.J. McPhaul, et al., Prediction of type 2 diabetes by hemoglobin A1c in two community-based cohorts, Diabetes Care 41 (2017) 60–68, https://doi.org/10.2337/dc17-0607.
[10] P.W.F. Wilson, J.B. Meigs, L. Sullivan, C.S. Fox, D.M. Nathan, RB Sr D'Agostino, Prediction of incident diabetes mellitus in middle-aged adults: the framingham offspring study, Arch. Intern. Med. 167 (2007) 1068–1074, https://doi.org/10.1001/archinte.167.10.1068.
[11] B.S. Wessler, Y.H.L. Lai, W. Kramer, M. Cangelosi, G. Raman, J.S. Lutz, et al., Clinical prediction models for cardiovascular disease, Circulation: Cardiovascular Quality and Outcomes 8 (2015) 368–375, https://doi.org/10.1161/CIRCOUTCOMES.115.001693.
[12] S. Singh, R. Shankar, G.P. Singh, Prevalence and associated risk factors of hypertension: a cross-sectional study in urban varanasi, Int. J. Hypertens. 2017 (2017) e5491838, https://doi.org/10.1155/2017/5491838.
[13] A. Floegel, N. Stefan, Z. Yu, K. Mühlenbruch, D. Drogan, H.-G. Joost, et al., Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach, Diabetes 62 (2013) 639–648, https://doi.org/10.2337/db12-0495.
[14] L. Sun, L. Liang, X. Gao, H. Zhang, P. Yao, Y. Hu, et al., Early prediction of developing type 2 diabetes by plasma acylcarnitines: a population-based study, Diabetes Care 39 (2016) 1563–1570, https://doi.org/10.2337/dc16-0232.
[15] X.S. Li, Z. Wang, T. Cajka, J.A. Buffa, I. Nemet, A.G. Hurd, et al., Untargeted metabolomics identifies trimethyllysine, a TMAO-producing nutrient precursor, as a predictor of incident cardiovascular disease risk, JCI Insight 3 (2018), https://doi.org/10.1172/jci.insight.99096.
[16] P. McGranaghan, A. Saxena, M. Rubens, J. Radenkovic, D. Bach, L. Schleußner, et al., Predictive value of metabolomic biomarkers for cardiovascular disease risk: a systematic review and meta-analysis, Biomarkers 25 (2020) 101–111, https://doi.org/10.1080/1354750X.2020.1716073.
[17] S. Dietrich, A. Floegel, C. Weikert, C. Prehn, J. Adamski, T. Pischon, et al., Identification of serum metabolites associated with incident hypertension in the European prospective investigation into cancer and nutrition–potsdam study, Hypertension 68 (2016) 471–477, https://doi.org/10.1161/HYPERTENSIONAHA.116.07292.
[18] N. Mars, J.T. Koskela, P. Ripatti, T.T.J. Kiiskinen, A.S. Havulinna, J.V. Lindbohm, et al., Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers, Nat. Med. 26 (2020) 549–557, https://doi.org/10.1038/s41591-020-0800-0.
[19] M.A. Simonson, A.G. Wills, M.C. Keller, M.B. McQueen, Recent methods for polygenic analysis of genome-wide data implicate an important effect of common variants on cardiovascular disease risk, BMC Med. Genet. 12 (2011) 146, https://doi.org/10.1186/1471-2350-12-146.
[20] A.V. Khera, M. Chaffin, K.G. Aragam, M.E. Haas, C. Roselli, S.H. Choi, et al., Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations, Nat. Genet. 50 (2018) 1219–1224, https://doi.org/10.1038/s41588-018-0183-z.
[21] J.W. O'Sullivan, S. Raghavan, C. Marquez-Luna, J.A. Luzum, S.M. Damrauer, E. A. Ashley, et al., Polygenic risk scores for cardiovascular disease: a scientific statement from the American heart association, Circulation 146 (2022) e93–e118, https://doi.org/10.1161/CIR.0000000000001077.
[22] E. Evangelou, H.R. Warren, D. Mosen-Ansorena, B. Mifsud, R. Pazoki, H. Gao, et al., Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits, Nat. Genet. 50 (2018) 1412–1425, https://doi.org/10.1038/s41588-018-0205-x.
[23] J. Elliott, B. Bodinier, T.A. Bond, M. Chadeau-Hyam, E. Evangelou, K.G. M. Moons, et al., Predictive accuracy of a polygenic risk score–enhanced prediction model vs a clinical risk score for coronary artery disease, JAMA 323 (2020) 636–645, https://doi.org/10.1001/jama.2019.22241.
[24] V. Lyssenko, A. Jonsson, P. Almgren, N. Pulizzi, B. Isomaa, T. Tuomi, et al., Clinical risk factors, DNA variants, and the development of type 2 diabetes, N. Engl. J. Med. 359 (2008) 2220–2232, https://doi.org/10.1056/NEJMoa0801869.
[25] Q.Q. Huang, N. Sallah, D. Dunca, B. Trivedi, K.A. Hunt, S. Hodgson, et al., Transferability of genetic loci and polygenic scores for cardiometabolic traits in British Pakistani and Bangladeshi individuals, Nat. Commun. 13 (2022) 4664, https://doi.org/10.1038/s41467-022-32095-5.
[26] C. Herder, M. Karakas, W. Koenig, Biomarkers for the prediction of type 2 diabetes and cardiovascular disease, Clin. Pharmacol. Therapeut. 90 (2011) 52–66, https://doi.org/10.1038/clpt.2011.93.
[27] Ruiz-Canela M, Hruby A, Clish CB, Liang L, Martínez-González MA, Hu FB. Comprehensive Metabolomic Profiling and Incident Cardiovascular Disease: A Systematic review. J. Am. Heart Assoc. n.d.;6:e005705. https://doi.org/10.1161/JAHA.117.005705.
[28] J.L. Vassy, J.B. Meigs, Is Genetic testing useful to predict type 2 diabetes? Best Pract. Res. Clin. Endocrinol. Metabol. 26 (2012) 189–201, https://doi.org/10.1016/j.beem.2011.09.002.
[29] J.P.A. Ioannidis, N.A. Patsopoulos, E. Evangelou, Heterogeneity in meta-analyses of genome-wide association investigations, PLoS One 2 (2007) e841, https://doi.org/10.1371/journal.pone.0000841.
[30] J.S. Kaufman, L. Dolman, D. Rushani, R.S. Cooper, The contribution of genomic research to explaining racial disparities in cardiovascular disease: a systematic review, Am. J. Epidemiol. 181 (2015) 464–472, https://doi.org/10.1093/aje/kwu319.
[31] K. Luijken, R.H.H. Groenwold, B. Van Calster, E.W. Steyerberg, M. van Smeden, Impact of predictor measurement heterogeneity across settings on the performance of prediction models: a measurement error perspective, Stat. Med. 38 (2019) 3444–3459, https://doi.org/10.1002/sim.8183.
[32] M.Z.I. Chowdhury, T.C. Turin, Variable selection strategies and its importance in clinical prediction modelling, Family Medicine and Community Health 8 (2020) e000262, https://doi.org/10.1136/fmch-2019-000262.
[33] H.S. Rajula, G. Verlato, M. Manchia, N. Antonucci, V. Fanos, Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment, Medicina 56 (2020) 455, https://doi.org/10.3390/medicina56090455.

[34] S.F. Weng, J. Reps, J. Kai, J.M. Garibaldi, N. Qureshi, Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 12 (2017) e0174944 https://doi.org/10.1371/journal.pone.0174944.
[35] C. Krittanawong, H.U.H. Virk, S. Bangalore, Z. Wang, K.W. Johnson, R. Pinotti, et al., Machine learning prediction in cardiovascular diseases: a meta-analysis, Sci. Rep. 10 (2020) 16057, https://doi.org/10.1038/s41598-020-72685-1.
[36] K. Eichler, M.A. Puhan, J. Steurer, L.M. Bachmann, Prediction of first coronary events with the Framingham score: a systematic review, Am. Heart J. 153 (2007) 722–731.e8, https://doi.org/10.1016/j.ahj.2007.02.027.
[37] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda, Machine learning and data mining methods in diabetes research, Comput. Struct. Biotechnol. J. 15 (2017) 104–116, https://doi.org/10.1016/j.csbj.2016.12.005.
[38] K.D. Silva, W.K. Lee, A. Forbes, R.T. Demmer, C. Barton, J. Enticott, Use and performance of machine learning models for type 2 diabetes prediction in community settings: a systematic review and meta-analysis, Int. J. Med. Inf. 143 (2020) 104268, https://doi.org/10.1016/j.ijmedinf.2020.104268.
[39] E. Martinez-Ríos, L. Montesinos, M. Alfaro-Ponce, L. Pecchia, A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data, Biomed. Signal Process Control 68 (2021) 102813, https://doi.org/10.1016/j.bspc.2021.102813.
[40] G. Battineni, G.G. Sagaro, N. Chinatalapudi, F. Amenta, Applications of machine learning predictive models in the chronic disease diagnosis, J. Personalized Med. 10 (2020) 21, https://doi.org/10.3390/jpm10020021.
[41] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gøtzsche, J.P. Ioannidis, et al., The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, PLoS Med. 6 (7) (2009 Jul 21), 10.1371/journal.pmed.1000100.
[42] C. Schardt, M.B. Adams, T. Owens, S. Keitz, P. Fontelo, Utilization of the PICO framework to improve searching PubMed for clinical questions, BMC Med. Inf. Decis. Making 7 (2007) 16, https://doi.org/10.1186/1472-6947-7-16.
[43] S. Oliver, Systematic functional analysis of the yeast genome, Trends Biotechnol. 16 (9) (1998 Sept 1) 373–378, https://doi.org/10.1016/S0167-7799(98)01214-1.
[44] M. Ouzzani, H. Hammady, Z. Fedorowicz, A. Elmagarmid, Rayyan—a web and mobile app for systematic reviews, Syst. Rev. 5 (2016) 210, https://doi.org/10.1186/s13643-016-0384-4.
[45] G.A. Walford, B.C. Porneala, M. Dauriz, J.L. Vassy, S. Cheng, E.P. Rhee, et al., Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes, Diabetes Care 37 (2014) 2508–2514, https://doi.org/10.2337/dc14-0560.
[46] S.-J. Hahn, S. Kim, Y.S. Choi, J. Lee, J. Kang, Prediction of type 2 diabetes using genome-wide polygenic risk score and metabolic profiles: a machine learning analysis of population-based 10-year prospective cohort study, EBioMedicine 86 (2022), https://doi.org/10.1016/j.ebiom.2022.104383.
[47] J. Liu, L. Wang, Y. Qian, Q. Shen, M. Yang, Y. Dong, et al., Metabolic and genetic markers improve prediction of incident type 2 diabetes: a nested case-control study in Chinese, J. Clin. Endocrinol. Metabol. 107 (2022) 3120–3127, https://doi.org/10.1210/clinem/dgac487.
[48] P. Louca, T.Q.B. Tran, C du Toit, P. Christofidou, T.D. Spector, M. Mangino, et al., Machine learning integration of multimodal data identifies key features of blood pressure regulation, EBioMedicine 84 (2022), https://doi.org/10.1016/j.ebiom.2022.104243.
[49] K.G. Moons, J.A. de Groot, W. Bouwmeester, Y. Vergouwe, S. Mallett, D. G. Altman, et al., Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist, PLoS Med. 11 (2014), https://doi.org/10.1371/journal.pmed.1001744.
[50] M.J. Downes, M.L. Brennan, H.C. Williams, R.S. Dean, Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS), BMJ Open 6 (2016) e011458, https://doi.org/10.1136/bmjopen-2016-011458.
[51] Ottawa hospital research institute n.d. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. (Accessed 20 October 2023).
[52] A. Stang, Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses, Eur. J. Epidemiol. 25 (2010) 603–605, https://doi.org/10.1007/s10654-010-9491-z.
[53] J.A. Sterne, M.A. Hernán, B.C. Reeves, J. Savović, N.D. Berkman, M. Viswanathan, et al., ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions, BMJ 355 (2016) i4919, https://doi.org/10.1136/bmj.i4919.
[54] R.F. Wolff, K.G.M. Moons, R.D. Riley, P.F. Whiting, M. Westwood, G.S. Collins, et al., PROBAST: a tool to assess the risk of bias and applicability of Prediction model studies, Ann. Intern. Med. 170 (1) (2019) 51, https://doi.org/10.7326/M18-1376.
[55] G.S. Collins, J.B. Reitsma, D.G. Altman, K. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement, BMC Med. 13 (2015) 1, https://doi.org/10.1186/s12916-014-0241-z.
[56] S.M. Schüssler-Fiorenza Rose, K. Contrepois, K.J. Moneghetti, W. Zhou, T. Mishra, S. Mataraso, et al., A longitudinal big data approach for precision health, Nat. Med. 25 (2019) 792–804, https://doi.org/10.1038/s41591-019-0414-6.
[57] A. Wesolowska-Andersen, C.A. Brorsson, R. Bizzotto, A. Mari, A. Tura, R. Koivula, et al., Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: an IMI DIRECT study, Cell Reports Medicine 3 (2022) 100477, https://doi.org/10.1016/j.xcrm.2021.100477.
[58] LH van Dongen, P.P. Harms, M. Hoogendoorn, D.S. Zimmerman, E.M. Lodder, LM 't Hart, et al., Discovery of predictors of sudden cardiac arrest in diabetes: rationale and outline of the RESCUED (REcognition of Sudden Cardiac arrest vUlnErability in Diabetes) project, Open Heart 8 (2021) e001554, https://doi.org/10.1136/openhrt-2020-001554.
[59] G. Drouard, M. Ollikainen, J. Mykkänen, O. Raitakari, T. Lehtimäki, M. Kähönen, et al., Multi-omics integration in a twin cohort and predictive modeling of blood pressure values, OMICS A J. Integr. Biol. 26 (2022) 130–141, https://doi.org/10.1089/omi.2021.0201.
[60] J. Carrasco-Zanini, M. Pietzner, E. Wheeler, N.D. Kerrison, C. Langenberg, N. J. Wareham, Multi-omic prediction of incident type 2 diabetes, Diabetologia (2023), https://doi.org/10.1007/s00125-023-06027-x.
[61] L. Bouzit, S. Malspeis, J.A. Sparks, J. Cui, E.W. Karlson, K. Yoshida, et al., Assessing improved risk prediction of rheumatoid arthritis by environmental, genetic, and metabolomic factors, Semin. Arthritis Rheum. 51 (5) (2021) 1016–1022, https://doi.org/10.1016/j.semarthrit.2021.07.006.
[62] X. Wu, C.C. Wang, Y. Cao, J. Li, Z. Li, H. Ma, et al., Novel genetic risk and metabolic signatures of insulin signaling and androgenesis in the anovulation of polycystic ovary syndrome, Engineering 23 (2023) 103–111, https://doi.org/10.1016/j.eng.2022.08.013.
[63] N. Atabaki-Pasdar, M. Ohlsson, A. Viñuela, F. Frau, H. Pomares-Millan, M. Haid, et al., Predicting and elucidating the etiology of Fatty Liver Disease: a machine learning modeling and Validation Study in the IMI direct cohorts, PLoS Med. 17 (6) (2020 Jun 19), https://doi.org/10.1371/journal.pmed.1003149.
[64] M. Guasch-Ferré, A. Hruby, E. Toledo, C.B. Clish, M.A. Martínez-González, J. Salas-Salvadó, et al., Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis, Diabetes Care 39 (2016) 833–846, https://doi.org/10.2337/dc15-2251.
[65] J. Morze, C. Wittenbecher, L. Schwingshackl, A. Danielewicz, A. Rynkiewicz, F. B. Hu, et al., Metabolomics and type 2 diabetes risk: an updated systematic review and meta-analysis of prospective cohort studies, Diabetes Care 45 (2022) 1013–1024, https://doi.org/10.2337/dc21-1705.
[66] R.P. Igo, T.G. Kinzy, J.N.C. Bailey, Genetic risk scores, Curr Protoc Hum Genet 104 (2019) e95, https://doi.org/10.1002/cphg.95.
[67] American Diabetes Association, 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2021, Diabetes Care 44 (2020) S15–S33, https://doi.org/10.2337/dc21-S002.
[68] A.C. Schrimpe-Rutledge, S.G. Codreanu, S.D. Sherrod, J.A. McLean, Untargeted metabolomics strategies-challenges and emerging directions, J. Am. Soc. Mass Spectrom. 27 (2016) 1897–1905, https://doi.org/10.1007/s13361-016-1469-y.
[69] J. Dupuis, C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A.U. Jackson, et al., New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk, Nat. Genet. 42 (2010) 105–116, https://doi.org/10.1038/ng.520.
[70] B.F. Voight, L.J. Scott, V. Steinthorsdottir, A.P. Morris, C. Dina, R.P. Welch, et al., Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis, Nat. Genet. 42 (2010) 579–589, https://doi.org/10.1038/ng.609.
[71] A. Jonsson, C. Ladenvall, T.S. Ahluwalia, J. Kravic, U. Krus, J. Taneera, et al., Effects of common genetic variants associated with type 2 diabetes and glycemic traits on α- and β-cell function and insulin action in humans, Diabetes 62 (2013) 2978–2983, https://doi.org/10.2337/db12-1627.
[72] P. Würtz, V.-P. Mäkinen, P. Soininen, A.J. Kangas, T. Tukiainen, J. Kettunen, et al., Metabolic signatures of insulin resistance in 7,098 young adults, Diabetes 61 (2012) 1372–1380, https://doi.org/10.2337/db11-1355.
[73] N.D. Palmer, R.D. Stevens, P.A. Antinozzi, A. Anderson, R.N. Bergman, L. E. Wagenknecht, et al., Metabolomic profile associated with insulin resistance and conversion to diabetes in the insulin resistance atherosclerosis study, J. Clin. Endocrinol. Metabol. 100 (2015) E463–E468, https://doi.org/10.1210/jc.2014-2357.
[74] E.P. Rhee, S. Cheng, M.G. Larson, G.A. Walford, G.D. Lewis, E. McCabe, et al., Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans, J. Clin. Invest. 121 (2011) 1402–1411, https://doi.org/10.1172/JCI44442.
[75] B.L. Rodriguez, D.R. Labarthe, B. Huang, J. Lopez-Gomez, Rise of blood pressure with age. New evidence of population differences, Hypertension 24 (1994) 779–785, https://doi.org/10.1161/01.HYP.24.6.779.
[76] E. Pinto, Blood pressure and ageing, Postgrad. Med. 83 (976) (2007 Feb 1) 109–114, https://doi.org/10.1136/pgmj.2006.048371.
[77] G.C. Linderman, J. Lu, Y. Lu, X. Sun, W. Xu, K. Nasir, et al., Association of Body Mass index with blood pressure among 1.7 million Chinese adults, JAMA Netw. Open 1 (4) (2018 Aug 17), https://doi.org/10.1001/jamanetworkopen.2018.1271.
[78] E.R. Rademacher, D.R.J. Jacobs, A. Moran, J. Steinberger, R.J. Prineas, A. Sinaiko, Relation of blood pressure and body mass index during childhood to cardiovascular risk factor levels in young adults, J. Hypertens. 27 (2009) 1766, https://doi.org/10.1097/HJH.0b013e32832e8cfa.
[79] H.R. Warren, E. Evangelou, C.P. Cabrera, H. Gao, M. Ren, B. Mifsud, et al., Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk, Nat. Genet. 49 (3) (2017) 403–415, https://doi.org/10.1038/ng.3768.
[80] G.B. Ehret, M.J. Caulfield, Genes for blood pressure: an opportunity to understand hypertension, Eur. Heart J. 34 (13) (2013) 951–961, https://doi.org/10.1093/eurheartj/ehs455.
[81] Y. Zheng, B. Yu, D. Alexander, T.H. Mosley, G. Heiss, J.A. Nettleton, et al., Metabolomics and incident hypertension among blacks, Hypertension 62 (2013) 398–403, https://doi.org/10.1161/HYPERTENSIONAHA.113.01166.
[82] C. Menni, D. Graham, G. Kastenmüller, N.H.J. Alharbi, S.M. Alsanosi, M. McBride, et al., Metabolomic identification of a novel pathway of blood

pressure regulation involving hexadecanedioate, Hypertension 66 (2015) 422–429, https://doi.org/10.1161/HYPERTENSIONAHA.115.05544.
[83] R. Couronné, P. Probst, A.-L. Boulesteix, Random forest versus logistic regression: a large-scale benchmark experiment, BMC Bioinf. 19 (2018) 270, https://doi.org/10.1186/s12859-018-2264-5.
[84] M.A. Friedl, C.E. Brodley, Decision Tree Classification of land cover from remotely sensed data, Rem. Sens. Environ. 61 (3) (1997) 399–409, https://doi.org/10.1016/s0034-4257(97)00049-7. Sept.
[85] J.V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, J. Clin. Epidemiol. 49 (1996) 1225–1231, https://doi.org/10.1016/S0895-4356(96)00002-9.
[86] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004 Aug) 199–222, https://doi.org/10.1023/b:stco.0000035301.49549.88.
[87] J.C. Jakobsen, C. Gluud, J. Wetterslev, et al., When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts, BMC Med. Res. Methodol. 17 (2017) 162, https://doi.org/10.1186/s12874-017-0442-1.
[88] M. Chagnon, J. O'Loughlin, J.C. Engert, I. Karp, M.-P. Sylvestre, Missing single nucleotide polymorphisms in genetic risk scores: a simulation study, PLoS One 13 (2018), https://doi.org/10.1371/journal.pone.0200630.
[89] R. Wei, J. Wang, M. Su, E. Jia, S. Chen, T. Chen, et al., Missing value imputation approach for mass spectrometry-based Metabolomics Data, Sci. Rep. 8 (2018), https://doi.org/10.1038/s41598-017-19120-0.
[90] J. Antonelli, B.L. Claggett, M. Henglin, A. Kim, G. Ovsak, N. Kim, et al., Statistical workflow for feature selection in human Metabolomics Data, Metabolites 9 (2019) 143, https://doi.org/10.3390/metabo9070143.
[91] M. Feurer, F. Hutter, Hyperparameter optimization, in: F. Hutter, L. Kotthoff, J. Vanschoren (Eds.), Automated Machine Learning: Methods, Systems, Challenges, Springer International Publishing, Cham, 2019, pp. 3–33, https://doi.org/10.1007/978-3-030-05318-5_1.
[92] B. Efron, The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, 1982, https://doi.org/10.1137/1.9781611970319.
[93] T.-T. Wong, Performance evaluation of classification algorithms by K-fold and leave-one-out cross validation, Pattern Recogn. 48 (9) (2015) 2839–2846, https://doi.org/10.1016/j.patcog.2015.03.009. Sept.
[94] E.W. Steyerberg, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, Springer International Publishing, Cham, 2019, https://doi.org/10.1007/978-3-030-16399-0.
[95] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, pp. 1137–1143.
[96] S. Parvandeh, H.-W. Yeh, M.P. Paulus, B.A. McKinney, Consensus features nested cross-validation, Bioinformatics 36 (2020) 3093–3098, https://doi.org/10.1093/bioinformatics/btaa046.
[97] D. Krstajic, L.J. Buturovic, D.E. Leahy, S. Thomas, Cross-validation pitfalls when selecting and assessing regression and classification models, J. Cheminf. 6 (2014) 10, https://doi.org/10.1186/1758-2946-6-10.
[98] S.A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M.A. Riegler, P. Halvorsen, et al., On evaluation metrics for medical applications of Artificial Intelligence, Sci. Rep. 12 (2022), https://doi.org/10.1038/s41598-022-09954-8.
[99] B. Van Calster, D.J. McLernon, M. van Smeden, L. Wynants, E.W. Steyerberg, Calibration: the achilles heel of predictive analytics, BMC Med. 17 (1) (2019), https://doi.org/10.1186/s12916-019-1466-7.
[100] C. Mood, Logistic regression: why we cannot do what we think we can do, and what we can do about it, Eur. Socio Rev. 26 (2010) 67–82, https://doi.org/10.1093/esr/jcp006.
[101] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[102] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles. https://doi.org/10.48550/arXiv.1802.03888, 2019.
[103] O. Osibogun, O. Ogunmoroti, E.D. Michos, Polycystic ovary syndrome and cardiometabolic risk: opportunities for cardiovascular disease prevention, Trends Cardiovasc. Med. 30 (2020) 399–404, https://doi.org/10.1016/j.tcm.2019.08.010.
[104] S. Lim, M.-R. Taskinen, J. Borén, Crosstalk between nonalcoholic fatty liver disease and cardiometabolic syndrome, Obes. Rev. 20 (2019) 599–611, https://doi.org/10.1111/obr.12820.
[105] M.A. Connelly, E.G. Gruppen, J.D. Otvos, R.P.F. Dullaart, Inflammatory glycoproteins in cardiometabolic disorders, autoimmune diseases and cancer, Clin. Chim. Acta 459 (2016) 177–186, https://doi.org/10.1016/j.cca.2016.06.012.
[106] A.J. Vickers, B. van Calster, E.W. Steyerberg, A simple, step-by-step guide to interpreting decision curve analysis, Diagnostic and Prognostic Research 3 (2019), https://doi.org/10.1186/s41512-019-0064-7.