Review

Risk Prediction of Cardiovascular Events by Exploration of Molecular Data with Explainable Artificial Intelligence

Annie M. Westerlund [1,2,†] , **Johann S. Hawe** [1,†] , **Matthias Heinig** [2,3,*] **and Heribert Schunkert** [1,4,*]

1   Department of Cardiology, Deutsches Herzzentrum München, Technical University Munich, Lazarettstrasse 36, 80636 Munich, Germany; annie.westerlund@tum.de (A.M.W.); johann.hawe@tum.de (J.S.H.)
2   Institute of Computational Biology, HelmholtzZentrum München, Ingolstädter Landstrasse 1, 85764 Munich, Germany
3   Department of Informatics, Technical University Munich, Boltzmannstrasse 3, 85748 Garching, Germany
4   Deutsches Zentrum für Herz- und Kreislaufforschung (DZHK), Munich Heart Alliance, Biedersteiner Strasse 29, 80802 Munich, Germany
*   Correspondence: matthias.heinig@helmholtz-muenchen.de (M.H.); schunkert@dhm.mhn.de (H.S.)
†   These authors contributed equally to this work.

Abstract: Cardiovascular diseases (CVD) annually take almost 18 million lives worldwide. Most lethal events occur months or years after the initial presentation. Indeed, many patients experience repeated complications or require multiple interventions (recurrent events). Apart from affecting the individual, this leads to high medical costs for society. Personalized treatment strategies aiming at prediction and prevention of recurrent events rely on early diagnosis and precise prognosis. Complementing the traditional environmental and clinical risk factors, multi-omics data provide a holistic view of the patient and disease progression, enabling studies to probe novel angles in risk stratification. Specifically, predictive molecular markers allow insights into regulatory networks, pathways, and mechanisms underlying disease. Moreover, artificial intelligence (AI) represents a powerful, yet adaptive, framework able to recognize complex patterns in large-scale clinical and molecular data with the potential to improve risk prediction. Here, we review the most recent advances in risk prediction of recurrent cardiovascular events, and discuss the value of molecular data and biomarkers for understanding patient risk in a systems biology context. Finally, we introduce explainable AI which may improve clinical decision systems by making predictions transparent to the medical practitioner.

Keywords: cardiovascular disease; coronary artery disease; genomics; proteomics; multi-omics; biomarkers; molecular networks; machine learning; AI; explainable artificial intelligence

Academic Editors: Claudia Kusmic and Francesca Forini

1. Introduction

Cardiovascular disease (CVD) is the most common source of death worldwide, ending almost 18 million lives every year [1]. In this paper, we focus on CVD caused by atherosclerosis, leading to coronary artery disease (CAD), peripheral arterial disease (PAD), and cerebrovascular disease (CeVD). Among these, CAD is the most prevalent [2]. A patient diagnosed with CAD carries plaque formations in vital vessels supplying the heart. This reduces blood flow and provokes symptoms such as chest pains and myocardial infarction (MI).

Most fatal cardiovascular events, such as stroke or MI, happen within months or years after the patient has been diagnosed with CVD [3]. In fact, many patients suffer severe complications or are in need of repeated interventions (*recurrent events*), after their initial presentation. The events pose a great risk to the patient, and may be an economic burden to society. Being able to accurately predict and explain the risk for a patient to develop recurrent events over time is essential for current treatment protocols and is a key

task for future clinical decision systems. Such decision systems would greatly facilitate medical practitioners to formulate personalized treatment plans. Specifically, the ability to explain decision patterns behind predicted risk opens up possibilities for designing novel prevention and treatment strategies, including development of drugs targeting patient-specific biomarkers.

The initial clinical presentation of a patient with CVD may offer useful information for predicting the future course of the disease. Risk prediction of CVD, as well as recurrent events, is traditionally done on the basis of clinical risk factors derived from blood tests, lab measurements, or questionnaires. A spectrum of features, or risk factors, are integrated in risk scores which are the current gold standard for risk prediction. These features include age, sex, BMI, hypertension, presence of diabetes mellitus, smoking status, family history, and cholesterol levels [2] (see Tables 1–4).The GRACE score [4], for example, builds on risk factors such as age, systolic blood pressure, cardiac arrest at admission, and initial serum creatinine concentration to predict severe recurrent events in patients presenting with an acute coronary syndrome (ACS). Other scores, such as the REACH score [5] and SMART score [6], on the other hand, predict recurrent cardiovascular events in patients with established CVD on the basis of age, smoking status, diabetes, medical history, or carotid ultrasound findings. However, these clinical risk factors alone can only identify a subset of patients with CAD at risk for recurrent events [7]. Many recent studies therefore address this limitation in prediction performance by augmenting the set of clinical risk factors, for example by imaging data or established molecular biomarkers.

Imaging techniques such as cardiac magnetic resonance (MR) and coronary computed tomography angiography (CCTA) are noninvasive and common diagnostic tools. Features extracted from the images, for instance with image segmentation, may be used directly as phenotypes or additional risk factors to predict incident CVD or recurrent events [8–11]. Unfortunately, imaging is expensive and therefore of limited availability.

Neither clinical risk factors nor imaging data capture information about the molecular underpinnings of a disease. Indeed, knowledge of such molecular mechanisms may aid clinicians in giving a more precise diagnosis, which may ultimately allow individually tailored treatments. For example, the mRNA expression of 50 specific genes can be successfully used to classify breast cancer into prognostic subtypes [12]. To identify prognostic markers, long-term studies with individual follow-ups are necessary to retrospectively identify molecules predictive of recurrent events. Once identified, the prognostic value of the markers can be prospectively scrutinized. For diagnostic markers, it suffices to assess molecular markers upon manifestation of the disease. This may provide more accurate patient stratification, compared to current clinical practices utilizing information from lab values or images.

Genome-wide association studies (GWAS), which associate genetic variants on the DNA to a phenotypic trait, such as CVD, have become a popular and successful tool within the past decade [13,14]. This has enabled researchers to identify hundreds of heritable genetic risk loci harboring CAD-associated DNA-variants, which affect CAD via additive and non-additive effects [15–18]. Although polygenic risk scores derived from disease-associated genetic variants facilitate accurate estimates of the individual lifetime risk for incident CAD [16,19,20], they are less accurate in predicting risk as a function of time or predicting molecular mechanisms driving disease progression. In contrast, data such as gene expression or protein expression reflect the environment and lifestyle of an individual. Such expression data can be measured from biosamples with current high-throughput molecular experiments, providing functional data which can be linked to the genotypes assessed in GWAS. Omics data involve the measurement of the entirety of a set of molecules such as mRNAs (transcriptomics) or proteins (proteomics) in a biological system (also referred to as *omics layers*). Functional omics data complement the genotyping data of a patient, and thus yields more detailed mechanistic insights into cellular processes, enabling, for instance, the identification of causal disease genes. Moreover, personalized gene regulatory networks may reveal genes or pathways involved in disease progression

of a specific patient, and may thus allow for personalized development or prescription of drugs targeting particular genes or pathways. Specifically, novel drugs can then be developed to target an identified causal gene or pathway. Indeed, analyses involving omics data are promising enhancers of CVD risk prediction [21,22]. This further suggests an important role for multi-omics data, regulatory networks, and pathway analysis when predicting and understanding recurrent events.

Lately, a multitude of reviews have covered the topics of classical risk prediction models [23], artificial intelligence (AI) applied to recognize CVD in various forms of imaging data [24–26], or AI for the purpose of electrocardiogram-based prediction and drug discovery [26]. Additionally, reviews considering the molecular and genetic aspects of CVDs have been published in recent years, mostly focusing on describing the complex processes underlying disease [14,27,28]. These reviews consider incident CVD rather than recurrent events. In contrast, we here focus on the recent advances in risk prediction of recurrent cardiovascular events based on clinical and molecular data and discuss the use of AI as a tool for (1) predicting risk and (2) designing personalized treatment strategies. Because prediction models are often quantitatively evaluated with performance metrics, we provide a brief description of common metrics in Table 5. First, we summarize traditional risk prediction models based on clinical risk factors. We then transition to the molecular aspects of CVD and highlight the added value of biomarkers and networks to understand and predict patient risk. Finally, we describe how challenges of risk prediction can now be addressed with AI, and how explainable AI (XAI) may be an integral part of future transparent and reliable clinical decision systems which are based on both clinical and molecular data.

Table 1. Table of clinical datasets used to train traditional risk scores. LDL—Low-density lipid; HDL—High-density lipid; BMI—Body mass index; LV—Left ventricle; CKD—Chronic kidney disease.

Dataset	Cohort Size	Attributes	Follow Up	Ref.
American Framingham heart study	~15,000	**FRS attributes:** age, sex, diabetes, LDL and HDL cholesterol, smoking, systolic blood pressure	12 years	[29]
SCORE pooled dataset (12 European cohorts)	~21,000	Sex, smoking, total cholestrol, tot. chol./HDL ratio, systolic blood pressure	Average 13 years	[30]
ACC/AHA pooled cohort	~25,000	FRS attributes, tot. chol. treated/untreated systolic blood pressure	≥12 years	[31]
QRESEARCH (No diabetes or CVD at baseline)	~10 million	FRS attributes, social deprivation, family history, BMI, LV function, antihypertensive agent treatment	17 years (study length)	[32]
Suita dataset	~5600	FRS attributes, CKD	11.8 years	[33]

Table 2. Table of traditional risk scores based on clinical datasets. MI—myocardial infarction; TIA—transient ischaemic attack; CAD—coronary artery disease; CVD—cardiovascular disease; PH—proportional hazards.

Risk Score	Dataset	Clinical Question	Method	Performance	Ref.
Framingham risk score	American Framingham heart study	10-year risk; CAD, CVD events	Cox PH	0.733–0.841/ 0.769–0.847 (C-statistic, events average, male/female)	[34]
SCORE	SCORE pooled dataset	10-year risk; CAD/CVD mortality	Weibull	0.71–0.84 (AUC-ROC, EU countries)	[30]
ACC/AHA pooled cohort equations	ACC/AHA pooled cohort	10-year risk; atherosclerotic CVD events	Cox PH	0.713–0.818 (C-statistic, sex/race average)	[31]
QRISK	QRESEARCH	10-year risk; CVD (MI, CAD, stroke, TIA)	Cox PH	0.7674/0.7879 (AUC-ROC, male/female)	[35]
Suita	Suita dataset	10-year risk; CAD	Cox PH	0.835 (C-statistic)	[33]

Table 3. Table of clinical, genetic, and imaging datasets used to predict risk for recurrent events. CCTA—coronary computed tomography angiography; CVD—cardiovascular disease; CAD—coronary artery disease; CeVD—cerebrovascular disease; PAD—peripheral artery disease; ACS—acute coronary syndrome.

Dataset	Cohort Size	Type of Data	Baseline	Follow Up	Ref.
(**a**) GRACE registry	~102,000 (30 countries)	Clinical risk factors	ACS	6 months	[36]
(**b**) REACH registry	~68,000 (44 countries)	Clinical risk factors, demographics	CVD (CAD, CeVD, PAD)	1–2 years	[37]
(**c**) EuroAspire	IV/V: ~16,000 (27 countries)	Clinical risk factors, lifestyle	CAD	0.5–3 years	[38]
(**d**) TIMI	~8600	Clinical risk factors	ACS	2.5 years (median)	[39]
(**e**) CONFIRM registry	~50,000 (6 countries)	CCTA images, clinical risk factors	suspected CAD	2.3 years (median)	[40]
(**f**) UCC-SMART	~13,000	Clinical risk factors, carotid ultrasound, (567 patients with CCTA images	CVD	4.7 years (median)	[41]
(**g**) GENIUS-CHD	~186,000 (57 studies, 18 countries)	Genotype	CAD	1–15 years	[42]

Table 4. Table of clinical risk scores for predicting recurrent events. CVD—cardiovascular disease; MI—Myocardial infarction; PH—proportional hazards; CI—confidence interval.

Risk score	Dataset	Clinical Question	Method	Performance	Ref.
(**a**) GRACE	GRACE registry (43,810 patients, 14 countries)	6-month risk; death, or death/MI	Cox PH	0.7/0.82 (C-statistic, death/death-MI)	[4]
(**b**) REACH	REACH registry (49,689 patients, 44 countries)	20-month risk; CVD events, cardiovasc. death	Cox PH	0.67 [0.66, 0.68]/ 0.75 [0.73, 0.77] (C-statistic, 95% CI, CVD/death)	[5]
(**c**) EuroAspire	EuroAspire (IV/V, 27 countries, 12,484 patients)	2-year risk; CVD events or interventions	Weibull	0.67 [0.64, 0.70] (C-statistic, 95% CI)	[3]
(**d**) TRS2P	TIMI (8598 patients, 9 predictors)	3-year risk; Cardiovasc. death, MI, stroke	Cox PH	0.67 [0.65, 0.69] (C-statistic, 95% CI)	[39]
(**e**) CONFIRM	CONFIRM registry (20,300 patients)	2-year risk; death	Cox PH	0.682 (C-statistic)	[43]
(**f**) SMART	UCC-SMART (5788 patients, 14 predictors)	10-year risk; CVD events	Cox PH	0.68 [0.64, 0.71] (C-statistic, 95% CI)	[6]

Table 5. Description of performance metrics. Here, a "positive" label can, for example, correspond to "having CVD", while a "negative" may correspond to "healthy".

Metric	Description	Math. Definition
True positive (TP)	A positive sample correctly predicted by the model.	TP
True negative (TN)	A negative sample correctly predicted by the model.	TN
False positive (FP)	A sample wrongly classified as positive by the model.	FP
False negative (FN)	A sample wrongly classified as negative by the model.	FN
Precision	Fraction of true positives among the predicted positives.	$\dfrac{TP}{TP + FP}$
Recall (Sensitivity)	Fraction of positives that are correctly predicted.	$\dfrac{TP}{TP + FN}$
Specificity	Fraction of negatives that are correctly predicted.	$\dfrac{TN}{TN + FP}$
Accuracy	Fraction of correctly predicted positives and negatives.	$\dfrac{TN + TP}{TP + TN + FP + FN}$

Table 5. *Cont.*

Metric	Description	Math. Definition
ROC curve (Receiver Operating Characteristic)	A curve indicating performance of a classifier. The *Y*-axis shows recall and the *X*-axis shows s = (1-specificity)	$R(s)$
AUC-ROC (Area Under the Curve - ROC)	Quantitative performance measure based on ROC curve. Ranges from 0 to 1, where 1 corresponds to perfect, and 0.5 to random, classification.	$\int R(s)\mathrm{d}s$
C-statistic	Equivalent to AUC-ROC. Can be used for censored data (missing patient outcomes). r_i—predicted risk of patient i t_i—time to event, patient i $\varepsilon_i \in \{0, 1\}$ - whether (event) information exists. $1(x) = \{1 \text{ if } x, 0 \text{ otherwise}\}$	$\dfrac{\sum_{ij} \varepsilon_i 1(r_i > r_j) 1(t_i < t_j)}{\sum_{ij} \varepsilon_i 1(t_i < t_j)}$
PR curve (Precision Recall)	Similar to ROC curve. *Y*-axis shows precision and *X*-axis recall (r).	$P(r)$
AUC-PR (AUC—Precision-Recall)	Quantitative performance measure based on PR curve. Alternative to AUC-ROC.	$\int P(r)\mathrm{d}r$

2. Results

2.1. Traditional Risk Scores Based on Clinical and Imaging Data

2.1.1. Predicting the Risk of Incident CVD with Survival Models

The Framingham Risk Score (FRS) [2], Table 2 was the first modern risk prediction tool and is based on conventional clinical risk factors collected within the American Framingham heart study, Table 1. It was initially derived to predict the 10-year risk of CAD, but was later extended to predict a combination of CVD events: CAD, cerebrovascular disease, peripheral artery disease and heart failure [34]. Multiple studies have validated FRS on other cohorts [33,35,44]. The score has, for example, been found to overestimate the 10-year risk of CAD in a Japanese cohort [33], and the 10-year risk of CVD by 35% in a UK cohort [35].

The problem of FRS transferability has sparked the development of a multitude of risk scores using various regionally and ethnically diverse cohorts, as well as risk factor combinations. What all these scores have in common is the fact that the underlying historic data may overestimate event rates since better treatment modalities or improved environmental conditions may ameliorate risk over time. Tables 1 and 2 summarize some well-known and established traditional risk scores, as well as the datasets from which they were derived. Examples of traditional risk scores include the guideline-recommended pooled European SCORE [30] (Systematic COronary Risk Evaluation), the pooled cohort equations recommended by the ACC/AHA (American College of Cardiology/American Heart Association) [31], the ASSIGN score [45] from the Scottish Intercollegiate Guidelines Network (SIGN), the UK QRISK scores [35,46], and the Japanese Suita score [33]. The QRISK score, inspired by the ASSIGN score, makes use of social deprivation and family history as additional risk factors. This, together with risk factors such as body mass index (BMI) and left ventricle hypertrophy increased performance, as measured by the area under the receiver-operating characteristics curve (AUC-ROC) on the UK dataset compared to

ASSIGN and FRS. Later versions of the QRISK score are enhanced by additional risk factors related to atrial fibrillation, chronic kidney disease, migraine, and mental illness [46,47]. The NICE (National Institute for health and Care Excellence) guidelines of 2014, updated in 2016 [48], specifically recommend the QRISK2 score for estimating the 10-year risk of CVD. Inspired by QRISK2, the Suita score includes chronic kidney disease as an additional predictor, which improved the prediction of a 10-year risk for CAD on the Japanese cohort compared to FRS, according to the C-statistic (0.835 compared to 0.768).

2.1.2. Predicting Recurrent Events with Survival Models

One of the first models for predicting recurrent events is the Global Registry of Acute Coronary Events (GRACE) score [4], which has been recommended by the European Society of Cardiology (ESC) since 2014 [49], Tables 3a and 4a. It was designed to predict the 6-month risk of death and the combined risk of death or myocardial infarction (MI). According to the analysis, cardiac arrest upon admission and age were important predictors of death. The updated GRACE2.0 score [50] substituted uncommon attributes—such as the Killip class (to diuretic use within 24 h) and serum creatine concentration (to medical history of renal insufficiency)—in order to make the score more useful in practice.

Although the GRACE score has been successful, it is limited to predicting risk up to 6 months, while the risk of recurrent events can be elevated for several years after diagnosis. Models considering longer time-spans have therefore been developed in later years, including the models derived from the REduction of Atherothrombosis for Continued Health (REACH) registry [5], EuroAspire [3], TIMI [39] (Thrombolysis In Myocardial Infarction) and CONFIRM registry [9,43,51] (COronary CT angiography evaluatioN For clinical outcomes: an InteRnational Multicenter registry), Tables 3b–e and 4b–e. The models derived from the REACH registry [5] were trained to predict recurrent cardiovascular events and death over a time-span of 20 months in patients with established CVD, Table 4b. The dataset contained data from 44 countries, including individuals from the Middle East. An external validation study, however, found that the model underestimated the 20-month risk of recurrent cardiovascular events in individuals from the United Arab Emirates, and showed an overall performance C-statistic of 0.63 (0.06 standard error) [52]. In general, the overall difference in risk between patients of different nationalities may be related to different genetics, different lifestyles, and environmental factors, or even a different quality of healthcare. Therefore, a model trained on a multi-national dataset may exhibit varying performance when evaluated on different nationalities. These multi-national models may thus benefit from recalibration to specific regional cohorts before being applied in practice [53].

The first score predicting the 10-year risk of recurrent cardiovascular events in patients with CVD was developed using the UCC-SMART (Secondary Manifestations of ARTerial disease) cohort [41] from the Utrecht university in the Netherlands, Tables 3f and 4f. The dataset included clinical risk factors as well as carotid ultrasound findings. The SMART score [6], Table 4f, yielded a C-statistic of 0.683 (95% CI: [0.650, 0.717]), which slightly improved prediction compared to using the clinical risk factors (0.675, 95% CI: [0.642, 0.708]) or the carotid ultrasound findings (0.644, 95% CI: [0.609, 0.697]) alone [6]. The score is currently part of the ESC-recommended U-Prevent tool for predicting the 5- to 10-year risk of cardiovascular events when the patient has a history of CVD [53]. Still, recent studies have attempted to extend and improve the SMART score by including interventions as an end-point [54], and risk factors derived from CCTA images [8]. In the latter study [8], a 567-patient UCC-SMART sub-cohort with CCTA images was used to predict the 4-year risk of recurrent cardiovascular events or interventions. The additional CCTA-derived risk factors corresponded to coronary artery calcium, thoracic aortic calcium, and heart valve calcium. The findings showed that coronary artery calcium alone improved prediction accuracy from a C-statistic of 0.65 (95% CI: [0.59, 0.72]) without any calcification scores to 0.67 (95% CI: [0.61, 0.73]). By contrast, the thoracic and heart valve calcium did not improve risk prediction of recurrent events.

The added value of CCTA imaging risk factors has also been demonstrated by the CONFIRM risk score [43]. To build the risk score, the authors explored a mix of CCTA imaging and clinical risk factors to predict the 2-year risk of mortality for patients with suspected CAD, Tables 3e and 4e. Based on the ability of CCTA risk factors to predict death (measured by the C-statistic of single Cox models), the final model consisted of a pre-trained score and two manually extracted CCTA risk factors: the number of nearby segments with more than 50% stenosis, and the number of nearby segments with calcified or mixed plaques. The CONFIRM score increased prediction performance in terms of the C-statistic relative to scores such as FRS (CONFIRM—0.682; FRS—0.639), showing the added value of CCTA-derived attributes for predicting mortality. Lately, this risk score has been enhanced with AI for the risk prediction task, as well as an extended list of CCTA and clinical risk factors [9,51] (see Section 2.3). Regardless, the CONFIRM score and the recent SMART score with CCTA risk factors [8] suggest that information such as coronary artery calcium or segments with calcified plaques from CCTA images improve risk prediction of recurrent cardiovascular events, interventions, and death. Although CCTA imaging is a non-invasive alternative for diagnosis of CAD, images are currently not available for all patients, which limits the practical use of these scores. Moreover, the image-based risk scores mentioned here may benefit from AI-based segmentation as opposed to manual extraction for identifying and extracting risk factors from the images.

A direct comparison between models based on the original papers is tedious due to the highly diverse set of end-points, time-spans, and cohort characteristics. Most of the models presented here, however, have an online risk calculator [4,6], or present the model explicitly in the original paper [5,43]. Model availability is not only the first step towards usability in a clinical setting, but also paves the way for more in-depth benchmarking studies.

Complementing clinical and imaging risk factors, recent research has shown that molecular data can further enhance risk prediction of incident CVD and recurrent events [55–59] as additional risk factors. Moreover, 40–50% of the risk of CAD is heritable [60], making genotype data in particular a potentially important aspect for predicting and characterizing the risk of recurrent cardiovascular events. In the next section, we will describe those molecular aspects in more detail.

2.2. Molecular Aspects of Risk Prediction of Cardiovascular Events

2.2.1. Understanding CVD and Recurrent Events with Genotyping Data

Molecular data has been on the rise in recent years and is well-established in biomedical research. Figure 1 gives a brief overview on how molecular data can be used for the purpose of risk prediction of recurrent cardiovascular events. Genome-wide association studies (GWAS) have pinpointed hundreds of genetic loci and SNPs (Single Nucleotide Polymorphisms) contributing to the risk of developing CAD [14,16,61]. These SNPs also have value in predicting recurrent events as well as the efficacy of preventive treatments [17,62–66]. About a fifth of the CAD risk loci are located in the vicinity of genes related to CVD-relevant biological pathways, such as metabolism of lipoproteins or blood pressure [67]. The important roles of LDL cholesterol and hypertension for incident CVD events are well-established and therefore used in traditional models with clinical risk factors, Table 1. In line with this, current research has shown that lipoprotein(a) [68], or ceramides and phospholipids [69,70] in the blood pose good additional risk factors for risk prediction of recurrent events compared to the SMART and TRS2P scores in Chinese and Scandinavian CAD patients, respectively.
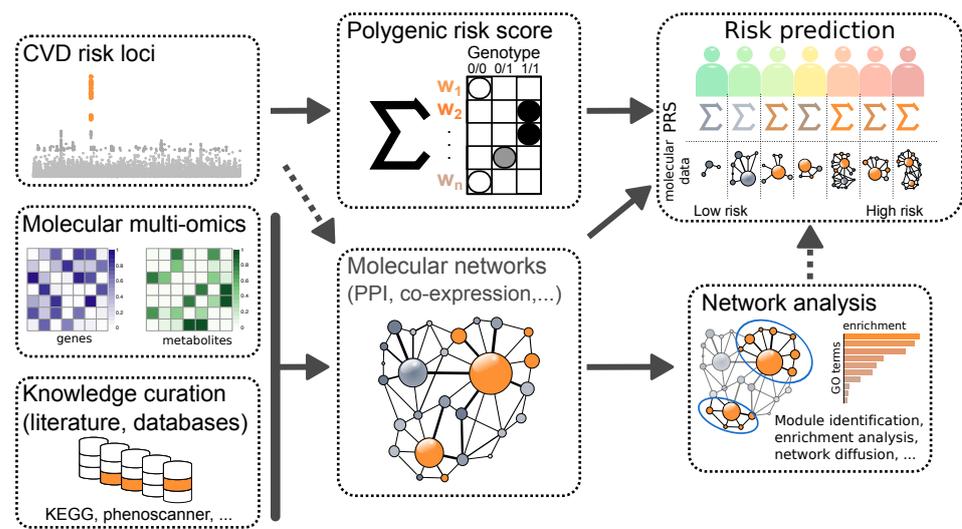
Figure 1. Overview of how molecular data can be used for understanding and predicting the risk of recurrent cardiovascular events. Genome-wide association studies (GWAS) can be used to identify CVD risk loci. Weights obtained from GWAS can be used to calculate a polygenic risk score. Moreover, the GWAS loci can be combined with multi-omics data and prior knowledge to construct regulatory networks. From these networks, it is possible to extract physiological pathways and network modules, as well as associate the level of activity of distinct network regions with high or low risk. The network information and polygenic risk score can be integrated together to improve risk prediction of recurrent cardiovascular events.

Studies based on genotype and conventional risk factor data from the GENIUS-CHD (GENetIcs of sUbSequent Coronary Heart Disease) consortium, Table 3g, showed that genetic variants associated with CAD may not necessarily be associated with recurrent events [71,72]. The known risk locus *9p21* [17,73], for example, did not show a strong association with recurrent acute CAD events [71]. This was demonstrated by the odds ratios from an age- and sex-adjusted meta-analysis of 49 studies, including almost 100,000 patients. In addition to this, the known CVD risk determinant *Factor V Leiden* was not associated with an increased risk of recurrent events, interventions or death in patients with CAD at baseline, as quantified by Cox hazard ratios [72]. The authors used 25 studies from GENIUS-CHD with a Factor-V Leiden genotype and conventional risk factors.

A disease-associated genetic locus, however, does not act in isolation, but rather through interactions with other genes, forming regulatory networks. So-called key-driver-genes have prominent positions in such networks as they, to some extent, determine the biological activity of a network [74,75]. Genetic variants modulating expression of key-driver-genes (eQTLs) come with a higher probability of affecting disease prevalence [28,76]. Unravelling the complex molecular architecture of gene regulatory networks giving rise to CAD is key to understanding disease progression, but may also be useful for predicting events. In general, some interactions between loci are assumed to be non-additive [18], necessitating the use of non-linear approaches such as AI models. However, so far, additive models have performed well in prediction tasks. For instance, additive models such as Polygenic Risk Scores (PRSs) (Figure 1) have so far been most successful to predict incident CVD [14,16,20,77,78], as well as recurrent events [79]. PRSs combine multiple genetic loci and are based on the fact that a combination of multiple variants of small effects can result in a larger risk for developing diseases like CAD, compared to specific rare monogenic mutations. The first genome-wide PRS for predicting CAD, among other diseases, was presented by Khera et al. in 2018 [16]. The results of this study have since then sparked new interest in PRS as a tool for predicting the genetic component of disease risk. Notably, a typical PRS consists of a two-step approach involving univariate variable selection through GWAS and subsequent multi-variate model fitting. The success of the

PRS then depends on these two steps. In a recent study, a LASSO (Least Absolute Shrinkage and Selection Operator)-based PRS was developed. This approach allowed for estimation and variable selection simultaneously, thus merging the two steps into one [80]. The PRS was successfully applied to UK biobank [81] data, demonstrating scalability as well as predictive performance in accord with other established PRSs.

Although CAD PRS has proven valuable as an additional risk factor in traditional survival models such as FRS for predicting CAD [55,82,83], the added value of CAD PRS for predicting recurrent events is debated. Early studies concluded that genetic scores of incident CAD (i.e., not trained to predict recurrent events) may not improve risk prediction of recurrent events [84], while more recent studies found that genetic scores of incident CAD can indeed lead to an improved prediction performance of recurrent events [85]. In the former study [84], 5742 patients from the UCC-SMART dataset were used to train two Cox regression models for predicting the 10-year risk of recurrent major cardiovascular events. The first model used the predictors of the original SMART model, while the second combined these predictors with a genetic score of incident CAD, including 30 variants associated with CAD. The added genetic score did not increase prediction performance, as measured by the C-statistic. By contrast, a genetic risk score based on 79 CAD risk variants was recently shown to slightly improve predictability of recurrent cardiovascular events within 5 years in a cohort of 1,667 patients with ACS at baseline. The CAD PRS alone yielded a C-statistic of 0.581 (95% CI: [0.566, 0.596]), while the model using only clinical data obtained 0.686 (95% CI: [0.630, 0.658]). Compared to this, the combination of clinical data and PRS gave a slightly increased C-statistic of 0.698 (95% CI: [0.640, 0.668]) [85]. Another study investigated associations between incident-CAD PRS and cardiovascular or fatal events, as well as conventional risk factors, in individuals with and without CAD at baseline [79]. Using 408,480 individuals from the UK biobank and genome-wide association studies (GWAS) summary data from the CARDIoGRAMplusC4D consortium, the analysis showed a larger association between high PRS and greater risk of cardiovascular events in healthy individuals compared to those with CAD at baseline. In contrast to the healthy controls, patients with CAD at baseline further displayed an inverse association between the CAD PRS and death or ischemic stroke [79]. The results in these studies [79,84,85] together with the studies on the GENIUS-CHD dataset [71,72] suggest that the CAD PRS may not necessarily be an optimal predictor of recurrent event risk. Instead, a potential improvement for prediction of recurrent events may be to directly build a PRS of recurrent events [65].

A PRS may capture the genetic risk for developing CVD or recurrent events, but it does not fully explain the molecular disease etiology of CVD. Correlations between SNPs, for example, make it difficult to pinpoint causal genes. Moreover, risk contributors such as environmental and lifestyle factors cannot be explained with genomics data alone. Therefore, since both environment and genetics influence the levels of transcripts, proteins, and metabolites, molecular data encode their combined effect on the state of a cell. Analysis of these data facilitates exposure of new aspects, including the underlying disease mechanism, and may allow for identification of new predictive biomarkers [56,86,87]. High levels of the high-sensitivity C-reactive protein (CRP), for example, are associated with a higher 10-year risk for recurrent MI or death in CAD and ACS patients who underwent percutaneous coronary intervention (PCI) [56]. Moreover, increased levels of the monocyte chemoattractant protein-1 (MCP-1) have been suggested to increase the risk of recurrent events in patients with CAD [87,88]. Hence, functional multi-omics data may provide important information that might be lacking in clinical risk factors and genotype data.

2.2.2. Integration of (Multi-)Omics Data

The possibility to generate readouts from different cellular information layers enables investigation of the molecular causes of a disease. Identified diagnostic biomarkers, for example, make it possible to guide treatment decisions and improve risk prediction. When collectively studied in a systems biology context, multi-omics have the potential to elucidate

complex molecular mechanisms and pathways driving a disease [27,89]. This information can then be used to propose novel prognostic and therapeutic targets.

Overall, quantitative molecular data aid researchers in the transition from genetic and clinical data to mechanistic studies. To this end, genotype information may be correlated with mRNA, protein, or metabolite levels to identify disease-associated variants affecting those traits (quantitative trait loci (QTL), Figure 1) [75,90–93]. The genotype of a genetic variant can, for example, be correlated with the mRNA expression levels of genes across a cohort of individuals to determine expression quantitative trait loci (eQTL).

In the context of CVD, studies identified significant differences between the mRNA levels obtained from the left ventricular myocardium of patients with dilated cardiomyopathy and non-diseased controls [93]. Here, the authors further integrated differentially expressed genes with genotyping information and published GWAS loci, yielding numerous functional candidate genes relevant to cardiac phenotypes. Integrative approaches, where the molecular data are combined with clinical data for diagnostic or prognostic purposes, have shown especially promising results in the past [56,86,87]. In the following paragraphs, we will describe how disease-specific biomarkers can be derived from patient-level (multi-)omics data. Moreover, we highlight the use of molecular interaction networks to elucidate disease mechanisms.

Biomarker discovery. In general, biomarker discovery requires at least two sets of biosamples: one from healthy individuals (controls) and one from diseased individuals (cases, for example, diagnosed CVD). The goal is to obtain molecules which allow prognosis or diagnosis of a particular disease through distinctive concentrations. Prognostic markers are found by assessing molecular features at baseline prior to developing the disease or an event. In a straightforward case, significant differences between disease carriers and healthy individuals can be identified directly from quantitative molecular readouts of, for example, mRNA, protein or metabolite levels. The top distinguishing markers can then be found by applying a specific p-value and fold-change cutoff. To better understand the biology involved, these markers can be further scrutinized in pathway analyses—including gene ontology (GO, [94]) or Kyoto Encyclopedia of Genes and Genomes (KEGG, [95]) pathway enrichment analyses—as visualized in Figure 1. This allows to pinpoint disease-specific biological processes. Although markers derived from these computational analyses have a huge potential for diagnosis and prognosis tasks, they also require proper experimental and clinical validation before reaching routine clinical practice [96].

In a recent study, miRNA levels in whole-blood samples from a 437-patient cohort were collected from patients undergoing cardiac catheterization [97]. These data were used in a differential expression analysis to identify miRNAs differently expressed between patient groups. The cohort consisted of CAD patients with or without recurrent events (mean follow-up 1.5 years), as well as healthy controls without angiographic CAD. The differential expression analysis identified distinct miRNA profiles for individuals at risk of recurrent myocardial infarction. Interestingly, they were able to distinguish individuals on standard medical therapy exhibiting recurrent events from such individuals without recurrent events on the basis of miRNA profiles, but not using clinical attributes. This indicates a large prognostic potential and an added predictive value of miRNA profiles in risk prediction of recurrent cardiovascular events.

In another study [98], the authors performed a differential expression analysis between the proteomes of 85 CAD patients exhibiting recurrent events in a 5-year follow-up versus 85 sex and age matched CAD patients without recurrent events in this time period. Biosamples from blood serum were taken at baseline in light of the prospective Indian Atherosclerosis Research Study (IARS). The analysis identified β-defensin-128 and histatin-3 as potential biomarker predictors of recurrent CAD events. These biomarkers together with conventional risk factors improved risk prediction of recurrent cardiovascular events compared to the basic model with only conventional risk factors (C-statistic of 0.8 compared to 0.677). This further shows the huge potential of using omics, and specifically proteomics, in clinical settings.

Network inference. Although differential expression analysis is a rather routine approach to obtain a list of biomarker candidates for understanding molecular disease processes, analyses involving molecular networks facilitate identification of more robust biomarkers and causal mechanisms [75,99–103]. By reverse-engineering disease-specific regulatory networks or pathways, for instance in a multi-omics setting, it is possible to trace the effects of a disease-associated genetic variant throughout the cell [104] (Figure 1). This could lead to a better understanding of disease mechanisms. Moreover, identification of network modules allows to pinpoint clusters of genes which are most relevant to a disease. This bears a large drugability potential by revealing genes that are neighboring CAD candidate genes in relevant target tissues [75,105].

A recent study investigated gene expression microarray data from CAD patients without a history of MI, patients with MI but without heart failure, and patients with MI and heart failure (46, 34, and 30 cases, respectively) [102]. After detecting differentially expressed genes and long-non-coding RNAs (lncRNAs) between the three groups, the authors applied GO and KEGG pathway enrichment analysis to annotate the identified genes with their biological context. By applying weighted gene co-expression network analysis (WGCNA) to infer disease-relevant regulatory networks, they, for instance, found FASN and DGKZ together with their co-expressed lncRNAs as novel genes associated with CAD progression.

Another interesting approach is the direct inference of differential gene regulatory networks for cases and controls. One study for example proposed a novel strategy to infer group-specific networks using curated prior knowledge on protein-protein-interactions (PPI) in a cancer context [101]. The authors were able to derive a prioritized list of genes from their networks for more accurate survival prediction on independent datasets. Although in a different disease context (cancer), this framework could potentially be adapted for prediction of recurrent cardiovascular events.

The work presented in [106] highlights another application of gene regulatory networks. In their work, the authors inferred 28 CAD-related regulatory networks in diverse tissues around a set of eSNPs (**e**xpression **S**ingle-**N**ucleotide **P**olymorphisms: genetic variants associated with gene expression levels). Through this, the authors identified novel genetic risk loci for CAD, which accounted for approximately 10% of heritability of CAD in addition to the 22% estimated by current GWAS studies. These new risk loci may be useful in the task of risk prediction by complementing current GWAS loci.

In another study, Franzén et al. utilized STAGE/STARNET genotype and gene expression data to investigate tissue-specific gene regulatory networks in CAD related tissues [103]. A step-wise approach was used to integrate cardiometabolic GWAS loci with tissue-specific expression quantitative trait loci (eQTL). The results were then fed into a causal inference pipeline to derive the gene regulatory networks. In particular, the results demonstrated that tissue-specific regulation of the *PCSK9* gene in visceral abdominal fat tissue, but not in other tissues, leads to an increased risk of elevated LDL/HDL ratios.

A different approach was implemented in a study by Sumathipala et al. [107]. The study curated a miRNA-protein-disease interaction network using several resources, such as miRTarBase [108], OMIM (Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD, USA), 30 July 2021. World Wide Web URL: https://omim.org/ accessed on 29 August 2021) [109] and GWAS databases through DisGeNet [110]. The curated multi-omics network with miRNA–protein, protein–protein and protein–disease interactions was used to predict novel miRNA–disease interactions and to identify disease–disease relationships as well as disease subtypes via a novel network diffusion method [107]. Their findings suggest that specific miRNAs could identify disease subtypes, implying that miRNA biomarkers can increase patient stratification accuracy and, potentially, risk prediction. Moreover, the authors were able to predict differentially expressed miRNAs based on network information which could enable drug target discovery. Because miRNAs are known to affect

CAD phenotypes [111], this work could be an important step towards improving clinical CAD models.

Although the methodological approaches mentioned here show great promise, most of them are not specifically applied to understand recurrent cardiovascular events. Added value could therefore be gained by exploring these approaches in such a context.

Pathway analysis. The goal of pathway analysis is to identify pathways or biological processes associated with CVD or recurrent events. The analysis may reveal known pathways and processes, such as cell adhesion, immune response, apoptosis, blood coagulation and inflammation [98,106]. Finding novel genes or pathways, on the other hand, may implicate potential therapeutic targets. In contrast to de-novo network and biomarker inference, one can rely on established pathway knowledge and known disease-relevant molecules and networks (Figure 1). For instance, if we observe many small but highly coordinated changes of gene expression in many genes of the same pathway, it is possible to estimate the combined effect of these genes. Moreover, one can identify disease association on a pathway level, which might be neglected when only analyzing individual genes. In fact, many of the studies referred to above which involve network inference rely on such knowledge to verify their biomarker candidates through enrichment analyses [97,102,106].

Overall, a large number of publicly available resources has been created for this purpose. Most of these resources—such as BioGrid [112], STRING [113], KEGG [95], and ConsensusPathDB [114]—focus on providing information about associations between molecular entities. Some, for example KEGG, also highlight associations with phenotypes or diseases, and provide a broader view of the biological context of genes. For a more complete overview on potential pathway resources, Bader et al. compiled a list of over 700 pathway databases in their Pathguide web resource (http://pathguide.org/, 30 July 2021) [115].

In a recent study [116], the authors inferred novel biomarker candidates with information provided by the GO, disease ontology (DO, [117]) and STRING databases, rather than validating the findings using GO or KEGG. They first curated a list of 608 potentially CAD-related genes from the CADgene database [118] and then followed a step-wise procedure to filter the list of genes and construct a protein-protein-interaction network. The network analysis pinpointed the single gene (EGFR) which is part of the ErbB signaling pathway. Specifically, increased levels of EGFR could lead to a 3-fold increase in CAD risk. The findings were validated via Western blotting in a CAD case-and-control cohort (N = 342/342).

With the advent of AI, new models have been developed for deriving polygenic risk scores, estimating pathway enrichment, and predicting risk based on molecular biomarkers or network structures. So far, most network-based methods rely on only a few omics layers or do not extend to a setting involving clinical and imaging data. Therefore, sophisticated AI may be useful for integrating all these data in novel prediction models. The next section will consider recent efforts where AI has been employed for predicting risk of CVD and recurrent events.

2.3. Exploiting Artificial Intelligence for Risk Prediction of Cardiovascular Events

Artificial intelligence (AI), including deep learning, has gained momentum in recent years and has been successfully applied to a wide range of medical applications [119–124]; a hot topic of many reviews [25,26,125]. In contrast to the traditional survival models, AI models can be trained to recognize complex and nonlinear patterns given a large set of data. Similar to Cox regression models, parameterized AI models may facilitate personalized treatments by assessing the patient-specific risk. AI models, however, often operate in a discrete fashion, predicting occurrence of an event, or risk at a specific time point, while traditional survival models predict the risk over years and are able to handle censored data.

Here, we will focus on the added value of AI models for predicting cardiovascular events. Although most contemporary research has focused on tailoring AI for risk prediction of CVD [44,126–128], we will also cover some recent models considering recurrent

events [9,129]. First, however, we will provide a brief introduction of common terminology and AI models.

2.3.1. Brief Introduction to AI

A machine learning, or AI, model is roughly either an *unsupervised* or a *supervised* method. Unsupervised methods are used to explore patterns in the data, for example by grouping observations according to some similarity criterion (i.e., cluster the data), without knowing about any groups beforehand. While unsupervised methods may be of great use to identify subgroups of patients with potentially different risk profiles, we will focus on supervised methods in this section. Supervised AI methods rely on labeled data: each sample is paired with a label. For example, an individual which could be described by clinical values or molecular measurements may be labeled as "healthy" or "diagnosed with CVD".

Figure 2A visualizes a typical AI workflow. Each patient is first described by a common set of features. A feature is a particular variable or patient attribute. In CVD risk prediction, a feature may correspond to a risk factor, molecular expression level, or smoking status, and so forth. The data are then divided into three sets: training, validation, and test. During the *learning* or *training* phase, the parameters of the model are optimized based on an objective function which compares labels predicted by the model to observed (known) labels. Ideally, the final model should yield predictions that are as similar to the observed labels as possible, without compromising model generalizability, that is, avoid overfitting the model on the training data. Most AI models have non-trainable parameters, so called hyperparameters. To avoid overfitting, such parameters may be selected in a data-driven fashion with cross-validation. The model performance is then evaluated on the validation set according to a performance metric, Figure 2. The model with the hyperparameters yielding the best validation-set performance is selected as the final model and evaluated on the independent test set. A well-performing model can later be used to obtain predictions of single patients which were not part of the original dataset. Based on such predictions, the medical practitioner can assess the need for medications or additional follow-up visits.

Methods trained to recognize discrete class labels (e.g., healthy/diseased) are called *classifiers*, while methods trained on continuous labels (e.g., BMI or time to event) are referred to as *regressors*. AI models come in different flavors, including tree-based methods such as random forests [130] or XGBoost (eXtreme Gradient Boosting tree) [131], and various forms of neural networks [125,132] as examples. Tree-based methods typically consist of decision trees, which are made up of nodes. At each node of a decision tree, new branches are formed by splitting the training samples along a single feature. The split at a node is done to decrease the label heterogeneity of samples as much as possible at the resulting branches. Both random forest and XGBoost are ensemble methods consisting of multiple decision tree classifiers. Such methods avoid overfitting by voting or averaging across the different classifiers. Two other common methods are the *k*-Nearest-Neighbour (*k*NN) and Naive Bayes classifiers. In *k*NN, a patient's risk status is predicted as the most common class label among the *k* samples which are closest to the patient in feature (patient attribute) space. Naive Bayes, on the other hand, is a model derived from probability theory. It assumes that all features are independent, conditionally on the class labels. Using this assumption, a sample (patient) can be associated with the most probable class label.

Neural networks attempt to mimic the activation patterns of connected neurons in a brain responding to input data [133]. The simplest form of a neural network is the so-called perceptron (Figure 2B). The perceptron is a binary classifier which returns a weighted sum of the input layer "neurons" parsed through an activation function. The activation function can for example be a simple linear mapping or, in the case of a binary classifier, a logistic sigmoid function. As such, the perceptron is simply another way to describe logistic regression.
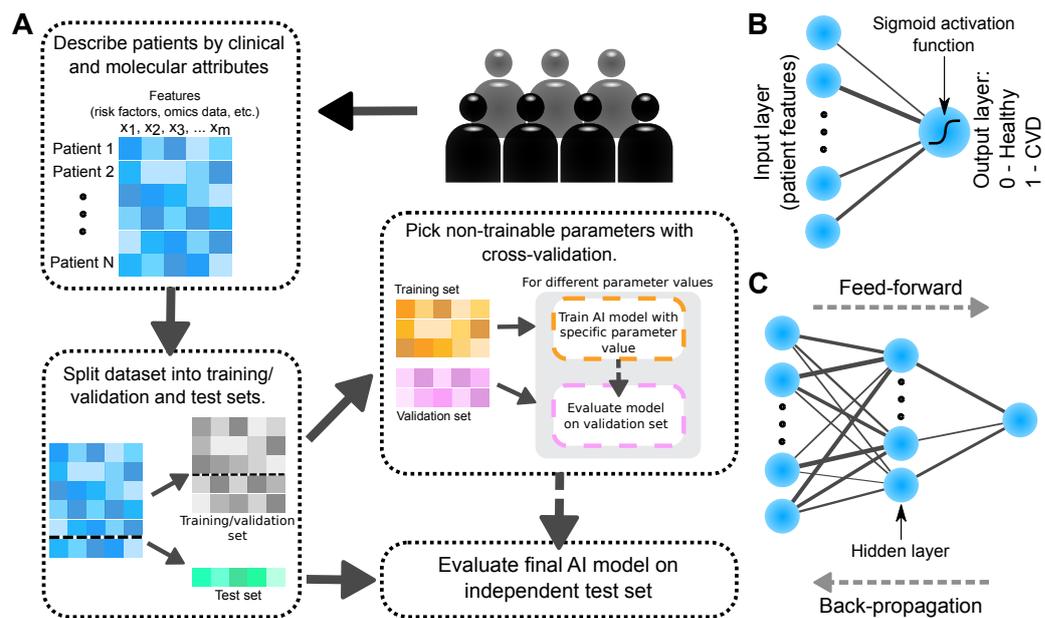
Figure 2. (**A**) Illustration of a typical AI workflow. Each patient is first described by the same set of numerical and/or categorical attributes (features), such as risk factors or gene expression levels. The data (patients) are then divided into training, validation and test sets. AI models with different values of non-trainable parameters (hyperparameters) are trained on the training set. The model performance is evaluated on the validation set according to some metric. A final model with the hyperparameters yielding the best validation-set performance is then evaluated on the independent test set. (**B**) Illustration of a perceptron neural network with a sigmoid activation function. (**C**) Illustration of a multi-layer neural network with one hidden layer. The arrows indicate direction of feed-forward and back-propagation passes.

By adding additional layers (*hidden layers*) with their own weights and activation functions between the input and output layer, the perceptron is extended to a multi-layer perceptron—the vanilla form of neural networks (Figure 2C). The extra hidden layers with their activation functions is one way to introduce the recognized complexity and non-linearity of neural networks [132]. The information of a sample, or patient input data, is then propagated from the input to the output layer via a number of hidden layers; a so-called *feed-forward pass*. The opposite flow of information (output-to-input layer) is called *back-propagation*. Back-propagation is especially useful for training the network by passing the objective function gradient with respect to the weights from the final output back to the first input layer in order to minimize prediction error. Furthermore, back-propagation can be used for explaining the network predictions (see Section 2.4).

In addition to the multi-layer perceptron, neural networks exist in other forms including recurrent neural networks which are able to model sequence data, autoencoders which can be used for dimensionality reduction, and Convolutional Neural Networks (CNNs) [132]. CNNs were originally developed for recognizing and segmenting objects in imaging data. Inspired by image processing, the weights are constrained to identical values across space and can be viewed as filters which are convolved with the input image at each layer. The constraint effectively reduces the number of parameters compared to fully connected neural networks. This allows more hidden layers, and thus more nonlinearity, to be introduced. In medical research, CNNs have been successfully applied to imaging datasets, for example to facilitate identification of malignant white blood cells [121,122], automatic cardiac and aortic image segmentation [11,123,134], and scoring of coronary artery calcium [135]. The use for AI in cardiovascular image analysis has been thoroughly reviewed elsewhere [136].

In general, neural networks are flexible in terms of architecture and can therefore be tailored to specific tasks. CNNs have for instance been successfully applied to 1D genome sequence data for predicting the functional impact of non-coding SNP-variants on epigenetic features such as transcription binding sites and DNA accessibility [137–139]. Consequently, CNNs may be also well-suited for the task of risk prediction of disease or disease progression based on genotype data. In a recent study, a CNN was specifically designed for genotype data and was shown to improve prediction accuracy of whether or not an individual will develop amyotrophic lateral sclerosis (ALS) compared to simpler AI models [140].

2.3.2. Utilizing Clinical and Imaging Data in AI Risk Prediction

Most studies so far have investigated the potential of AI for predicting incident disease rather than recurrent events, and typically make use of clinical risk factors. For example, an optimized AI pipeline—in terms of data imputation scheme, feature selection, and prediction model—was used to predict CVD in patients using 473 clinical data features from 423,604 patients from the UK biobank [44]. The so-called *autoprognosis* framework specifically selected this pipeline from a large pool of methods using Bayesian optimization. The selected and trained pipeline achieved an AUC-ROC of 0.774 (95% CI: [0.768, 0.780]), and thus outperformed both a Cox PH (proportional hazard) model fitted on the UK Biobank cohort using the conventional risk factors of the Framingham score (AUC-ROC 0.734 (95% CI: [0.729, 0.739]), Table 1) and a Cox PH model trained on the full set of attributes (AUC-ROC 0.758 (95% CI: [0.753, 0.763])), Table 6a.

In a study on prediction of presence of CAD, the performance of a CNN framework was compared to basic AI models, such as logistic regression, support-vector machine (SVM), and random forest [141], Table 6b. The models were trained on roughly 37,000 individuals described by 36 features from clinical data as well as demographic, laboratory and examination data. Although the findings could be strengthened statistically with for example confidence intervals, the results indicated that a CNN may perform similar to an SVM and random forest in terms of AUC-ROC and recall (0.768/0.773 compared to 0.774/0.776 and 0.7644/0.763), but slightly higher in terms of specificity (0.818 compared to 0.779 and 0.761). In another project, the added value of AI was investigated for the purpose of predicting the 5-year risk of CVD in a multi-ethnic cohort from Northern California [126], Table 6c. The ACC/AHA pooled cohort equations (PCE) [31] (Table 2) were compared to logistic regression, random forest, XGBoost and gradient boosted machine (GBM). Most AI models performed at least as well as the ACC/AHA pooled cohort equations in a patient-reduced dataset, Table 2. The three best models, the GBM, logistic regression with LASSO (L_1) regularization, and XGBoost, reached an AUC-ROC of 0.779 (95% CI: [0.760, 790]), 0.784 [0.765, 0.802] and 0.784 [0.766, 0.803] on the reduced dataset, while the ACC/AHA PCE reached 0.775 [0.755, 0.794] and logistic regression with L_2-regularization 0.749 [0.729, 0.770]. Although the simpler AI models did not necessarily improve performance, they allowed for a wider range of values in the input data and missing values than the ACC/AHA PCE. This allowed further improvement of the models in terms of AUC-ROC when trained on the full dataset, Table 6c.

In contrast to primary event prediction, another study derived three AI models (called the PRAISE models) for predicting the 1-year occurrence of recurrent MI, bleeding and death in patients with acute coronary syndrome (ACS) [129], Table 6d. Each model was constructed by selecting the best-performing classifier from a pool of four classifiers (adaptive boosting, naive Bayes, *k*NN, random forest). Model selection was performed using a validation set comprising 20% of the dataset. The models were trained on 25 features obtained from pooled international cohorts with clinical data of patients from Europe, Asia, North America, and South America. The final three models (adaptive boosting) achieved an average accuracy of 0.81–0.92 on the training dataset, as well as on the external (European) validation dataset) [129], Table 6.

The traditional CONFIRM model has lately been enhanced using a logit-boost model trained on a subset of approximately 10,000 patients from the CONFIRM dataset. The specific subset included extended CCTA imaging and clinical risk factors and longer follow-up time than the original model. The logit-boost model, which was trained to predict the 5-year risk of death in patients with suspected CAD, achieved an AUC-ROC of 0.79 (95% CI: [0.77, 0.81]). Thus, it outperformed traditional models such as the FRS (0.61 [0.59, 0.64]) and three CCTA-based scores (AUC-ROC 95% CI in the span [0.60, 0.66]) [9], Table 6e.

Table 6. Performance of AI models for predicting primary or recurrent events compared to traditional survival models. CVD—cardiovascular disease; CAD—coronary artery disease; ACS—acute coronary syndrome; PH—proportional hazards; CI—confidence interval; NPHS—National Pulmonary Hypertension Service; EHR—Electronic health records; LR—Logistic regression; RF—Random forest; MLP—multi-layer perceptron; NB—Naive Bayes.

Risk Score/ Method	Dataset	Clinical Question	Prediction Performance	Comparison (Cox PH)	Ref.
(**a**) Auto-prognosis framework	UK Biobank (clinical data, 423,604 patients)	5 year-risk; Fatal or non-fatal CVD event	*AUC-ROC, 95% CI:* 0.774 [0.768, 0.780]	*All attributes:* 0.758 [0.753, 0.763] *FRS attributes:* 0.734 [0.729, 0.739]	[44]
(**b**) CAD CNN (**1**) CNN (**2**) LR (**3**) SVM (**4**) RF (**5**) AdaBoost (**6**) MLP	NHANES (clinial, lab, demographic data, 37,079 patients)	Predict presence of CAD	*AUC-ROC:* (**1**) 76.87 (**2**) 71.29 (**3**) 77.64 (**4**) 76.24 (**5**) 71.63 (**6**) 72.61	-	[141]
(**c**) ASC AI (**1**) GBM (**2**) LR–L_1 (**3**) XGBoost (**4**) RF (**5**) LR-L_2	EHR (clinical data, socioecomics 262,923/ 131,721 patients)	5-year risk; MI, stroke, or fatal CAD	*AUC-ROC, 95% CI: (full/reduced)* (**1**) 0.835 / 0.779 [0.825, 0.846]/ [0.760, 0.790] (**2**) 0.784/0.825 [0.765, 0.802]/ [0.812, 0.839] (**3**) 0.784/0.830 [0.766, 0.803]/ [0.816, 0.843] (**4**) 0.773/0.831 [0.760,0.793]/ [0.820,0.842] (**5**) 0.749/0.808 [0.729,0.770]/ [0.795,0.820]	*ACC/AHA: (full/reduced)* - /0.775 [-, -]/ [0.755, 0.794]	[126]
(**d**) PRAISE (AdaBoost)	BleeMACS, RENAMI (clinical data, 19,826 patients)	1-year risk; (**1**) Death, (**2**) MI, (**3**) bleeding (ACS at baseline)	*AUC-ROC, 95% CI:* (**1**) 0.82 [0.78, 0.85] (**2**) 0.74 [0.70, 0.78] (**3**) 0.70 [0.66,0.75]	-	[129]
(**e**) CONFIRM (logit-boost)	CONFIRM registry (10,030 patients)	5-year risk; Death (suspected CAD at baseline)	*AUC-ROC, 95% CI:* 0.79 [0.77, 0.81]	*FRS:* 0.61 [0.59, 0.64]	[9]

Table 6. *Cont.*

Risk Score/ Method	Dataset	Clinical Question	Prediction Performance	Comparison (Cox PH)	Ref.
(f) 4D-survival (survival autoencoder)	NPHS (MR imaging, clinical data, 302 patients)	Survival-times *(pulmonary hypertension at baseline)*	*C-statistic, 95% CI:* 0.75 [0.70, 0.79]	*Clinical/imaging:* 0.64 [0.57, 0.70]	[11]
(g) CAD PRS **(1)** PRS **(2)** SVM **(3)** NB **(4)** RF **(5)** XGBoost	GerMIFS I-V, LURIC ($\sim$ 2.8M SNPs, 15,709 patients)	Genetic risk; CAD	*AUC-ROC, 95% CI:* **(1)** 0.92 [0.90, 0.94] **(2)** 0.82 [0.80, 0.85] **(3)** 0.82 [0.79, 0.84] **(4)** 0.75 [0.72, 0.78] **(5)** 0.74 [0.71, 0.77]	-	[20]

AI should in theory be able to model nonlinear data and therefore be superior to linear methods, such as the Cox model, for learning complex patterns. It can, however, be computationally expensive and sometimes tedious to select and train a model. The PRAISE models [129], together with the project by Ward et al. [126], suggest that ensemble models perform better than simpler AI models—including logistic regression, decision trees, *k*NN and naive Bayes—when predicting cardiovascular events and death, Table 6c–e. This was also suggested by the findings in a recent study predicting CVD in patients with type-2 diabetes [128]. The study, which was based on 172 CVD patients and 172 healthy controls, compared six models, namely naive Bayes, decision tree, random forest, SVM, logistic regression and *k*NN classifiers. Among these, random forest achieved the highest AUC-ROC (0.83 compared to 0.75–0.81 of the other models) [128]. In contrast, Dutta et al. found that the adaptive boosting classifier may perform similar to logistic regression, while the random forest and SVM had similar performance compared to the more advanced CNN when predicting the presence of CAD [141], Table 6b. The diverse results with respect to model performances may in part be attributed to insufficient choices of hyperparameters, thus highlighting some difficulties in finding optimal AI models. In general, simple models require less choices of hyperparameters and architectures, but are limited in complexity and flexibility.

In a clinical cohort, the follow-up data of some patients may be lost over time. This could be the case if no event happened within the time-frame of the study or if a patient drops out of a study by, for example, moving to a different region before exhibiting an event. The data are then *right-censored* at the time when the follow-up was lost. This type of data is difficult to directly make use of in basic AI models. Some AI models, however, allow for incorporation of survival models into their framework. Decision trees, and thus also random forest, can for example be adapted to right-censored data and be applied for risk prediction of cardiovascular events [142,143]. Moreover, the linear parametric function of the traditional Cox model can be replaced by a nonlinear function—for example, a neural network—which can improve the predictive performance compared to the original linear model [144,145]. Recently, an autoencoder neural network coupled with Cox regression was applied to sequences of cardiac magnetic resonance images to predict survival of 302 patients with pulmonary hypertension at baseline [11]. The model, which was trained on features from the CNN-segmented images, demonstrated an improved C-statistic (0.75, 95% CI: [0.70, 0.79]), compared to a traditional Cox regression model on clinical and imaging risk factors (0.64, 95% CI: [0.57, 0.70]), Table 6f.

2.3.3. Utilizing Molecular Data in AI Models

The use for AI to enhance CAD polygenic risk scores (PRSs) is an intriguing research question. A CAD PRS based on random forest preprocessing was recently developed on the GerMIFS I-V and LURIC datasets, Table 6g [20]. The dataset initially contained $\sim$2.8 million SNPs after quality control, which were reduced to $\sim$98,000 by removing highly correlated

SNPs. The variants were then filtered down to ~50,000 using the importance of SNPs from a trained random forest prior to computing the PRS (see Section 2.4 on explainable AI). Only SNPs with nonzero importance were used for predicting the genetic risk of CAD. Interestingly, an additive PRS model on the filtered set of SNPs proved superior to AI models such as random forest, naive Bayes, SVM, and XGBoost, Table 6. This indicated that a linear additive PRS model may be enough to predict CAD and that increasing model complexity with AI may be of little advantage. Lately, this PRS has been benchmarked against the earlier genome-wide PRS by Khera et al. [16] on the MONICA/KORA dataset, showing similar performances [83]. In addition, both these PRSs increased performance of CAD risk prediction when added to conventional risk factors, demonstrating the utility of genetics in risk prediction [83].

Networks based on co-expression have traditionally been inferred solely from gene expression data and simple correlation measures. This, however, often yields highly connected networks owing to spurious correlations. Improvements of these traditional methods are therefore more restrictive in terms of the number of edges in the networks. Restrictions are enforced through regularization-based approaches involving partial correlations [146,147]. Other extensions aim at combining data from multiple omics layers. This way, the resulting regulatory networks span multiple cellular information layers [148–150]. Currently, new avenues for network inference are being explored in disease contexts. Graph neural networks employ deep learning to derive molecular network structures and disease-associated genes [151]. These neural networks can then be used for inference and prediction tasks. In prediction tasks, information about expression levels and molecular interactions are used to prioritize disease genes and identify patient sub-types, for example, patients at low or high risk of recurrent events.

With graph-CNNs, it is now possible to use the molecular networks directly as input to the risk prediction models [152]. The novel approach in [152] integrated multi-omics data to pinpoint new omics signatures and biomarkers for disease using a combination of neural networks and a View Correlation Discovery Network (VCDN). This method was successfully applied in several disease contexts, including Alzheimer's disease and different types of cancers. In addition, graph-CNNs based on molecular data have been successful in predicting cancer genes [153,154] and cancer subtypes [155]. However, work is still needed to fully migrate this work into the field of CVD and recurrent events.

Multi-view neural networks have recently been proposed for analyses involving multi-omics data, including risk prediction [89]. This is carried out using a neural network with multiple input branches, such that each view corresponds to a specific type of omics data. The individual views then complement each other, ultimately leading to more comprehensive descriptions of the complex biological system at hand. Specifically, multi-view approaches allow to infer interactions between variables both within and between individual views [89]. As an example, multi-view models have been shown to outperform single modality-based approaches in predicting which patients with local memory loss symptoms will develop Alzheimer's disease (AD) [156]. Hence, these models may also be useful in the context of risk prediction of recurrent cardiovascular events, especially when different types of patient data are available such as clinical, multi-omics and imaging data.

2.4. Explaining Decisions Made by AI Models

Machine learning or artificial intelligence (AI) models are traditionally described as "black boxes" where the patterns in the input data giving rise to certain outputs are unknown. In other words, an algorithm may be able to achieve a near to perfect performance in a classification task, but it might not be possible for humans to understand the intrinsic factors driving the classification. The problem with lacking insight into model decisions appears when a model learns to predict accurately based on irrelevant features and thus generalizes poorly to other datasets. A recent study, for instance, reported that an accurate AI model trained to identify COVID-19 in chest radiographs actually failed to make use of the relevant information in the images [157]. Due to a consistent patient

positioning during imaging, the model instead recognized COVID-19-positive patients based on their shoulder regions. Unfortunately, the black-box behaviour of AI models is especially attributed to non-linear and complex models which are particularly suitable for advanced risk prediction.

Explainable AI (XAI) is a collection of methods which reveal the decision patterns of these complex models [158]. XAI may unveil novel aspects of a dataset and the disease itself by disclosing the predictive *importance* (or *relevance*) of input features, such as risk factors, attributes and protein expression levels. Simple models—e.g., principal component analysis (PCA), logistic regression and the Cox regression model—provide feature importance directly by their model parameters. Another simple model is the decision tree, which reveals decision patterns in the tree structure (see also Section 2.3.1). The so-called Gini feature importance is weighted by how much the impurity (label heterogeneity) of the data are decreased by a node split [159]. Such feature analyses provide an average importance over all samples (patients). This can often be extended to multiple classes by training models to distinguish each class from the rest, yielding insights related to, for instance, specific groups of patients [160]. However, the real power of explainable AI for clinical purposes comes with methods allowing to explain decisions made on individual patients. Figure 3 shows the main idea of using XAI in clinical decision systems. After the AI model has performed a patient-specific prediction, the medical practitioner can ask why the patient was associated with a certain risk, thus linking the model output to a specific set of clinical, imaging, or molecular attributes. This not only reveals the relevant risk factors, but may also be used to provide insights to the molecular mechanisms underlying disease by exploiting pathway databases and estimating gene set enrichment. All this information can guide the clinician when deciding on a treatment strategy.
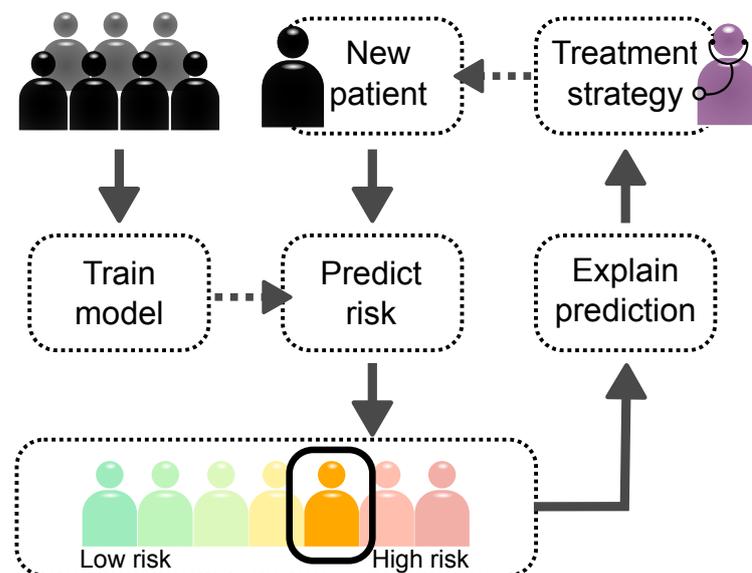

Figure 3. A typical clinical decision system with explainable AI. An AI model is first trained on a cohort containing for example clinical, imaging or multi-omics data. The trained model is then used to predict the risk for a patient to develop the disease or specific symptoms. Finally, the explainable AI provides information about the decision patterns, which helps the medical practitioner to assess faithfulness of the prediction and formulate a treatment strategy.

Moreover, by knowing the decision pattern, the clinician would be able to assess faithfulness of the prediction. For this reason, explainability is suggested to be an ethical requirement for future clinical decision systems [157]. Here, we introduce different explainable AI methods and give a brief overview of applications in medicine and analysis of molecular pathways.

2.4.1. Model-Specific Relevance Explanations

Explainable AI represents a large branch within the machine learning community. Most XAI methods are designed for a specific AI model, with lots of efforts put to deciphering neural networks. Within the group of model-specific methods, we for instance find the Gini importance for random forest classifiers [159], interpretable local surrogates [161], which estimate importance by approximation with a simpler model, and methods based on the architectural and back-propagation properties of neural networks [158,162].

Gini importance has found successful applications in pathway analyses [99,143,163]. An advantage of using random forests as opposed to the traditional univariate statistical testing is the implicit incorporation of interaction information, such as gene–gene interactions. In a random forest-based pathway approach, one classifier or regressor is trained per pathway category. These approaches use genes as predictor variables and estimate variable importance for each gene and pathway category. The distribution of gene relevances are then statistically compared between the genes of a specific pathway category and a subset of the background genes to test for pathway enrichment [143,163]. A recent benchmark study [99] found that random forest-based approaches outperformed traditional statistical methods on simulated data. The Gini importance provides an average variable importance over the sample population, which may help to disclose overall molecular drivers of the disease. With patient-specific XAI, on the other hand, it would be possible to obtain information about the pathways driving disease in each individual and thus possible to decide which drugs might be effective to treat that patient.

A patient-specific XAI method is the layer-wise relevance propagation (LRP), which is specific to neural networks. LRP operates via simple back-propagation rules which distribute relevance between layers [158]. The approach is thus able to disentangle otherwise highly complex and non-linear networks one layer at a time. The Deep Taylor Decomposition (DTD) framework specifically applies a Taylor decomposition at each layer while propagating relevance from the output layer to the input layer [162]. DTD can be viewed as a theoretical framework which encompasses multiple back-propagation relevance methods, including LRP [158]. Although DTD and LRP are able to provide feature relevance on a patient-specific level, they can only be applied to specific models, and an overall implementation supporting all neural network topologies is still missing.

2.4.2. Model-Agnostic Relevance Explanations

A model-agnostic framework does not rely on a specific functional form or gradient and can therefore be applied to different types of AI models. A popular method for inferring feature predictability in supervised models is the permutation method, which calculates the increased prediction error after permuting the feature data. Although this can be applied to any AI model, a notable problem concerns correlated features. Specifically, two important yet correlated features may not affect prediction performance if not both features' data are permuted. The permutation method is simple to implement and use but does not yield patient-specific relevance explanations.

The Shapley Additive Explanations (SHAP) [164], based on the game-theoretical Shapley value [165], is both model-agnostic and patient-specific. Intuitively, the Shapley value considers players in a game and tries to estimate the contribution of a player to the game outcome. In a risk prediction context, a game could correspond to a patient, a player to a risk factor, and a game outcome to the risk of developing disease or events. The contribution of a player can then be obtained by comparing game outcomes in the presence and absence of the player. This difference can be computed on a per-game (patient-specific) basis, making SHAP interesting for clinical decision systems. SHAP was recently presented and developed for tree models, such as XGBoost [164]. The theoretical framework, however, can be extended to any other model, including neural networks.

2.4.3. Clinical Applications of XAI

XAI in clinical applications is still in its infancy, but LRP has already proven useful for interpreting CNNs used for image-based identification of MS [166], malignant leukocytes [122], and Alzheimer's disease [167]. Moreover, LRP has been adapted for graph-CNNs and used to identify and characterize molecular signatures related to cancer [153,154].

Concerning model-agnostic XAI, the permutation method was used to obtain the patient-averaged feature importance from the PRAISE models predicting recurrent events [129], Table 6d. The analysis showed that left-ventricle ejection fraction, age, hemoglobin level and diabetes were important for predicting recurrent cardiovascular events and death within 1 year of being diagnosed with ACS. Another example is found in the original SHAP paper [164]. The authors used the NHANES dataset of 14,407 patients, 79 features, including clinical, demographic and lab attributes, and 20 year follow up to predict mortality risk and applied SHAP to further analyze the data and trained models. Their analysis revealed previously unknown interactions between features, such as age and sex, for mortality prediction. The data specifically suggested that the largest difference in mortality risk between men and women is observed at the age of 60 years. The large difference was suggested to be a result of the high risk of cardiovascular mortality for men relative to women.

Altogether, XAI represents an intriguing avenue which, together with the flexibility of AI models, may greatly improve contemporary risk prediction models of recurrent cardiovascular events.

3. Discussion

Accurate risk prediction models are crucial for facilitating early prognosis and prevention of recurrent cardiovascular events, as well as for formulating personalized treatment strategies. Smart decision systems in clinical settings would therefore not only save human lives, but also reduce expenses related to follow up visits and interventions, alleviating this large burden on the individual as well as the society.

Given the recent studies highlighted in this review, risk prediction of recurrent cardiovascular events can be improved in two ways in the future: (1) by using better-quality datasets with both clinical and molecular attributes, and (2) by using advanced, yet explainable, artificial intelligence (AI) models. Patient care could be improved drastically by leveraging the data already gathered in routine clinical work and combining them with additional molecular data. Therefore, the rise of multi-omics data in clinical research and the increased digitization of medical facilities is especially promising. However, the prospects of recurrent-event risk prediction in the context of CVD require additional efforts where molecular and clinical data are brought together into a unified framework. Moreover, clinical data are not always available directly to researchers, and technical and data privacy issues need to be overcome to successfully use them in medical research.

A recent comprehensive benchmark of 11 AI models and 18 datasets showed that, in a cancer context, integration of multi-omics and clinical data in an AI framework can sometimes improve prediction of survival time compared to the traditional Cox model based on clinical attributes alone [86]. The added value of omics data was especially clear in models which weighed or treated the omics layers and clinical data differently in a block-like fashion. Novel AI models with flexible architectures may therefore be particularly advantageous when handling datasets with a mix of traditional clinical risk factors, imaging, and multi-omics data. However, the benchmark also showed that whether or not a specific AI model leveraging multi-omics data actually improved prediction compared to a Cox model trained on clinical data alone depended on the specific dataset at hand. In the context of recurrent cardiovascular event risk prediction, the PRAISE models based on adaptive boosting [129] displayed a high AUC-ROC (Table 6d) and good calibration, but were not validated or compared to traditional scores such as the GRACE [50] or TIMI [39] scores (Table 4a,d). Moreover, although the CONFIRM logit-boost model (Table 6e) outperformed traditional models trained on less attributes [9], it was

not compared to a Cox model trained on the same attributes as the logit-boost model. Thus, more work is needed to properly benchmark the predictive ability of simpler AI models, ensemble models, neural networks and the Cox model when predicting recurrent cardiovascular events. A proper benchmark of the different traditional and AI models listed in this review would require one or, preferably, multiple datasets with multi-national individuals. The dataset(s) should be big enough to enable thorough statistical analysis of national- or event-specific model predictability. Moreover, they should not only contain the correct data of all risk factors used in each model, but also information about the different event outcomes of interest. Although most datasets can be used to evaluate a basic model, such as the Framingham risk score, they typically cannot be used to evaluate models with uncommon attributes—such as risk factors extracted from imaging data—due to missing data. Therefore, a completely unbiased comparison of these models on different datasets is difficult.

Although successful in predicting incident CAD, the use of known genetic CAD risk loci for predicting recurrent events is still under investigation. For instance, the application of polygenic risk scores (PRS) for CAD, while beneficial for predicting incident disease, does not necessarily enhance prediction of recurrent events. This issue could, however, be alleviated through systematic evaluation and application of state-of-the-art network and pathway modelling. Numerous resources containing both functional omics data and known interactions between molecular entities have been established for this purpose. Current developments in artificial intelligence and machine learning, such as graph neural networks, could advance recurrent risk prediction by properly utilizing these resources. As discussed in this review, novel biomarkers predicting risk could be implicated by combining clinical risk factors with disease-associated genetic variants, additional molecular data and with established knowledge about molecular interactions.

Finally, clinical decision systems developed in future studies should be able to explain their predictions on a patient-level basis. This would inform the medical practitioner about prediction reliability, and aid the formulation of a personalized treatment plan. In particular, explainable AI may reveal the functional disease mechanisms on a molecular level when multi-omics data together with pathway databases are included in the decision system framework. Being able to integrate the different layers and types of data via a comprehensive explanatory AI framework may thus render a more holistic view of the patient and their risk for recurrent cardiovascular events. New developments in the AI field need to take these factors into account in order to mature from research to clinical applications.

Abbreviations

The following abbreviations are used in this manuscript:

CVD	Cardiovascular disease
CAD	Coronary artery disease
PAD	Peripheral artery disease
CeVD	Cerebrovascular disease
MR	magnetic resonance
CCTA	Coronary computed tomography angiography
GWAS	Genome-wide association studies
AI	Artificial intelligence
XAI	Explainable artificial intelligence
FRS	Framingham risk score
SCORE	Systematic COronary Risk Evaluation
ACC/AHA	American College of Cardiology/American Heart Association
TIMI	Thrombolysis In Myocardial Infarction
GRACE	Global Registry of Acute Coronary Events
SMART	Secondary Manifestations of ARTerial disease
REACH	REduction of Atherothrombosis for Continued Health
CONFIRM	COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter registry
GENIUS-CHD	Genetics of Subsequent Coronary Heart Disease
PRS	Polygenic risk score
LASSO	Least Absolute Shrinkage and Selection Operator
CNN	Convolutional neural networks
XGBoost	eXtreme Gradient Boosting

References

1. Ruan, Y.; Guo, Y.; Zheng, Y.; Huang, Z.; Sun, S.; Kowal, P.; Shi, Y.; Wu, F. Cardiovascular disease (CVD) and associated risk factors among older adults in six low-and middle-income countries: Results from SAGE Wave 1. *BMC Public Health* **2018**, *18*, 778. [CrossRef]
2. Wilson, P.W.F.; D'Agostino Ralph, B.; Levy, D.; Belanger Albert, M.; Silbershatz, H.; Kannel William, B. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* **1998**, *97*, 1837–1847. [CrossRef] [PubMed]
3. De Bacquer, D.; Ueda, P.; Reiner, Z.; De Sutter, J.; De Smedt, D.; Lovic, D.; Gotcheva, N.; Fras, Z.; Pogosova, N.; Mirrakhimov, E.; et al. Prediction of recurrent event in patients with coronary heart disease: The EUROASPIRE Risk Model: Results from a prospective study in 27 countries in the WHO European region—The EURObservational Research Programme (EORP) of the European Society of Cardiology (ESC). *Eur. J. Prev. Cardiol.* **2020**, zwaa128. [CrossRef]
4. Fox, K.A.A.; Dabbous, O.H.; Goldberg, R.J.; Pieper, K.S.; Eagle, K.A.; Werf, F.V.D.; Avezum, A.; Goodman, S.G.; Flather, M.D.; Anderson, F.A.; et al. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: Prospective multinational observational study (GRACE). *BMJ* **2006**, *333*, 1091. [CrossRef] [PubMed]
5. Wilson, P.W.F.; D'Agostino, R.; Bhatt, D.L.; Eagle, K.; Pencina, M.J.; Smith, S.C.; Alberts, M.J.; Dallongeville, J.; Goto, S.; Hirsch, A.T.; et al. An International Model to Predict Recurrent Cardiovascular Disease. *Am. J. Med.* **2012**, *125*, 695–703.e1. [CrossRef]
6. Dorresteijn, J.A.N.; Visseren, F.L.J.; Wassink, A.M.J.; Gondrie, M.J.A.; Steyerberg, E.W.; Ridker, P.M.; Cook, N.R.; van der Graaf, Y.; SMART Study Group. Development and validation of a prediction rule for recurrent vascular events based on a cohort study of patients with arterial disease: The SMART risk score. *Heart* **2013**, *99*, 866–872. [CrossRef]
7. Fernández-Friera, L.; Fuster, V.; López-Melgar, B.; Oliva, B.; García-Ruiz, J.M.; Mendiguren, J.; Bueno, H.; Pocock, S.; Ibáñez, B.; Fernández-Ortiz, A.; et al. Normal LDL-Cholesterol Levels Are Associated With Subclinical Atherosclerosis in the Absence of Risk Factors. *J. Am. Coll. Cardiol.* **2017**, *70*, 2979–2991. [CrossRef] [PubMed]
8. van't Klooster, C.C.; van der Graaf, Y.; Nathoe, H.M.; Bots, M.L.; de Borst, G.J.; Visseren, F.L.J.; Leiner, T.; Asselbergs, F.W.; Nathoe, H.M.; de Borst, G.J.; et al. Added value of cardiovascular calcifications for prediction of recurrent cardiovascular events and cardiovascular interventions in patients with established cardiovascular disease. *Int. J. Cardiovasc. Imaging* **2021**, *37*, 2051–2061. [CrossRef] [PubMed]
9. Motwani, M.; Dey, D.; Berman, D.S.; Germano, G.; Achenbach, S.; Al-Mallah, M.H.; Andreini, D.; Budoff, M.J.; Cademartiri, F.; Callister, T.Q.; et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *Eur. Heart J.* **2017**, *38*, 500–507. [CrossRef] [PubMed]
10. Meyer, H.V.; Dawes, T.J.W.; Serrani, M.; Bai, W.; Tokarczuk, P.; Cai, J.; de Marvao, A.; Henry, A.; Lumbers, R.T.; Gierten, J.; et al. Genetic and functional insights into the fractal structure of the heart. *Nature* **2020**, *584*, 589–594. [CrossRef]
11. Bello, G.A.; Dawes, T.J.W.; Duan, J.; Biffi, C.; de Marvao, A.; Howard, L.S.G.E.; Gibbs, J.S.R.; Wilkins, M.R.; Cook, S.A.; Rueckert, D.; et al. Deep-learning cardiac motion analysis for human survival prediction. *Nat. Mach. Intell.* **2019**, *1*, 95–104. [CrossRef]

12. Parker, J.S.; Mullins, M.; Cheang, M.C.; Leung, S.; Voduc, D.; Vickery, T.; Davies, S.; Fauron, C.; He, X.; Hu, Z.; et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J. Clin. Oncol.* **2009**, *27*, 1160–1167. [CrossRef] [PubMed]

13. Visscher, P.M.; Wray, N.R.; Zhang, Q.; Sklar, P.; McCarthy, M.I.; Brown, M.A.; Yang, J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **2017**, *101*, 5–22. [CrossRef] [PubMed]

14. Chen, Z.; Schunkert, H. Genetics of coronary artery disease in the post GWAS era. *J. Intern. Med.* **2021**, 1–13 [CrossRef]

15. Wünnemann, F.; Sin Lo, K.; Langford-Avelar, A.; Busseuil, D.; Dubé, M.P.; Tardif, J.C.; Lettre, G. Validation of Genome-Wide Polygenic Risk Scores for Coronary Artery Disease in French Canadians. *Circ. Genom. Precis. Med.* **2019**, *12*, e002481. [CrossRef]

16. Khera, A.V.; Chaffin, M.; Aragam, K.G.; Haas, M.E.; Roselli, C.; Choi, S.H.; Natarajan, P.; Lander, E.S.; Lubitz, S.A.; Ellinor, P.T.; et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **2018**, *50*, 1219–1224. [CrossRef]

17. Aragam, K.G.; Jiang, T.; Goel, A.; Kanoni, S.; Wolford, B.N.; Weeks, E.M.; Wang, M.; Hindy, G.; Zhou, W.; Grace, C.; et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *medRxiv* **2021**. [CrossRef]

18. Zeng, L.; Moser, S.; Mirza-Schreiber, N.; Lamina, C.; Coassin, S.; Nelson, C.P.; Annilo, T.; Franzén, O.; Kleber, M.E.; Mack, S.; et al. Cis-epistasis at the LPA locus and risk of cardiovascular diseases. *Cardiovasc. Res.* **2021**, cvab136. [CrossRef] [PubMed]

19. Koyama, S.; Ito, K.; Terao, C.; Akiyama, M.; Horikoshi, M.; Momozawa, Y.; Matsunaga, H.; Ieki, H.; Ozaki, K.; Onouchi, Y.; et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **2020**, *52*, 1169–1177. [CrossRef]

20. Gola, D.; Erdmann, J.; Müller-Myhsok, B.; Schunkert, H.; König, I.R. Polygenic risk scores outperform machine learning methods in predicting coronary artery disease status. *Genet. Epidemiol.* **2020**, *44*, 125–138. [CrossRef]

21. Ruiz-Canela, M.; Hruby, A.; Clish, C.B.; Liang, L.; Martínez-González, M.A.; Hu, F.B. Comprehensive Metabolomic Profiling and Incident Cardiovascular Disease: A Systematic Review. *J. Am. Heart Assoc.* **2017**, *6*, e005705. [CrossRef]

22. Kessler, T.; Schunkert, H. Coronary Artery Disease Genetics Enlightened by Genome-Wide Association Studies. *JACC Basic Transl. Sci.* **2021**, *6*, 610–623. [CrossRef]

23. Damen, J.A.A.G.; Hooft, L.; Schuit, E.; Debray, T.P.A.; Collins, G.S.; Tzoulaki, I.; Lassale, C.M.; Siontis, G.C.M.; Chiocchia, V.; Roberts, C.; et al. Prediction models for cardiovascular disease risk in the general population: Systematic review. *BMJ* **2016**, *353*, i2416. [CrossRef] [PubMed]

24. Martin-Isla, C.; Campello, V.M.; Izquierdo, C.; Raisi-Estabragh, Z.; Baessler, B.; Petersen, S.E.; Lekadir, K. Image-Based Cardiac Diagnosis With Machine Learning: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 1. [CrossRef] [PubMed]

25. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises. *Proc. IEEE* **2021**, *109*, 820–838. [CrossRef]

26. Mathur, P.; Srivastava, S.; Xu, X.; Mehta, J.L. Artificial Intelligence, Machine Learning, and Cardiovascular Disease. *Clin. Med. Insights Cardiol.* **2020**, *14*, 1179546820927404. [CrossRef]

27. Doran, S.; Arif, M.; Lam, S.; Bayraktar, A.; Turkez, H.; Uhlen, M.; Boren, J.; Mardinoglu, A. Multi-omics approaches for revealing the complexity of cardiovascular disease. *Briefings Bioinform.* **2021**, *22*, bbab061. [CrossRef]

28. Heinig, M. Using Gene Expression to Annotate Cardiovascular GWAS Loci. *Front. Cardiovasc. Med.* **2018**, *5*, 59. [CrossRef]

29. Anderson, K.M.; Wilson, P.W.; Odell, P.M.; Kannel, W.B. An updated coronary risk profile. A statement for health professionals. *Circulation* **1991**, *83*, 356–362. [CrossRef] [PubMed]

30. Conroy, R.; Pyörälä, K.; Fitzgerald, A.; Sans, S.; Menotti, A.; De Backer, G.; De Bacquer, D.; Ducimetière, P.; Jousilahti, P.; Keil, U.; et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *Eur. Heart J.* **2003**, *24*, 987–1003. [CrossRef]

31. Goff, D.C.; Lloyd-Jones, D.M.; Bennett, G.; Coady, S.; D'Agostino, R.B.; Gibbons, R.; Greenland, P.; Lackland, D.T.; Levy, D.; O'Donnell, C.J.; et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation* **2014**, *129*, S49–S73. [CrossRef]

32. Hippisley-Cox, J.; Stables, D.; Pringle, M. QRESEARCH: A new general practice database for research. *J. Innov. Health Inform.* **2004**, *12*, 49–50. [CrossRef] [PubMed]

33. Nishimura, K.; Okamura, T.; Watanabe, M.; Nakai, M.; Takegami, M.; Higashiyama, A.; Kokubo, Y.; Okayama, A.; Miyamoto, Y. Predicting Coronary Heart Disease Using Risk Factor Categories for a Japanese Urban Population, and Comparison with the Framingham Risk Score: The Suita Study. *J. Atheroscler. Thromb.* **2014**, *21*, 784–798. [CrossRef] [PubMed]

34. D'Agostino, R.B.; Vasan Ramachandran, S.; Pencina Michael, J.; Wolf Philip, A.; Cobain, M.; Massaro Joseph, M.; Kannel William, B. General Cardiovascular Risk Profile for Use in Primary Care. *Circulation* **2008**, *117*, 743–753. [CrossRef]

35. Hippisley-Cox, J.; Coupland, C.; Vinogradova, Y.; Robson, J.; May, M.; Brindle, P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: Prospective open cohort study. *BMJ* **2007**, *335*, 136. [CrossRef] [PubMed]

36. Investigators, G. Rationale and design of the GRACE (Global Registry of Acute Coronary Events) Project: A multinational registry of patients hospitalized with acute coronary syndromes. *Am. Heart J.* **2001**, *141*, 190–199. [CrossRef]

37. Ohman, E.M.; Bhatt, D.L.; Steg, P.G.; Goto, S.; Hirsch, A.T.; Liau, C.S.; Mas, J.L.; Richard, A.J.; Röther, J.; Wilson, P.W.F.; et al. The REduction of Atherothrombosis for Continued Health (REACH) Registry: An international, prospective, observational

investigation in subjects at risk for atherothrombotic events-study design. *Am. Heart J.* **2006**, *151*, 786.e1–786.e10. [CrossRef] [PubMed]

38. Kotseva, K. The EUROASPIRE surveys: Lessons learned in cardiovascular disease prevention. *Cardiovasc. Diagn. Ther.* **2017**, *7*, 633–639. [CrossRef]

39. Bohula, E.A.; Bonaca, M.P.; Braunwald, E.; Aylward, P.E.; Corbalan, R.; De Ferrari, G.M.; He, P.; Lewis, B.S.; Merlini, P.A.; Murphy, S.A.; et al. Atherothrombotic Risk Stratification and the Efficacy and Safety of Vorapaxar in Patients With Stable Ischemic Heart Disease and Previous Myocardial Infarction. *Circulation* **2016**, *134*, 304–313. [CrossRef]

40. Min, J.K.; Dunning, A.; Lin, F.Y.; Achenbach, S.; Al-Mallah, M.H.; Berman, D.S.; Budoff, M.J.; Cademartiri, F.; Callister, T.Q.; Chang, H.J.; et al. Rationale and design of the CONFIRM (COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter) Registry. *J. Cardiovasc. Comput. Tomogr.* **2011**, *5*, 84–92. [CrossRef]

41. Simons, P.C.; Algra, A.; van de Laak, M.F.; Grobbee, D.E.; van der Graaf, Y. Second manifestations of ARTerial disease (SMART) study: Rationale and design. *Eur. J. Epidemiol.* **1999**, *15*, 773–781. [CrossRef]

42. Patel, R.S.; Tragante, V.; Schmidt, A.F.; McCubrey, R.O.; Holmes, M.V.; Howe Laurence, J.; Kenan, D.; Axel, Å.; Karin, L.; Virani Salim, S.; et al. Subsequent Event Risk in Individuals with Established Coronary Heart Disease. *Circ. Genom. Precis. Med.* **2019**, *12*, e002470. [CrossRef] [PubMed]

43. Hadamitzky, M.; Achenbach, S.; Al-Mallah, M.; Berman, D.; Budoff, M.; Cademartiri, F.; Callister, T.; Chang, H.J.; Cheng, V.; Chinnaiyan, K.; et al. Optimized Prognostic Score for Coronary Computed Tomographic Angiography: Results From the CONFIRM Registry (COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter Registry). *J. Am. Coll. Cardiol.* **2013**, *62*, 468–476. [CrossRef] [PubMed]

44. Alaa, A.M.; Bolton, T.; Angelantonio, E.D.; Rudd, J.H.F.; van der Schaar, M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS ONE* **2019**, *14*, e0213653. [CrossRef] [PubMed]

45. Woodward, M.; Brindle, P.; Tunstall-Pedoe, H.; SIGN Group on Risk Estimation. Adding social deprivation and family history to cardiovascular risk assessment: The ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* **2007**, *93*, 172–176. [CrossRef] [PubMed]

46. Hippisley-Cox, J.; Coupland, C.; Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ* **2017**, *357*, j2099. [CrossRef] [PubMed]

47. Collins, G.S.; Altman, D.G. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: Independent and external validation of an updated version of QRISK2. *BMJ* **2012**, *344*, e4181. [CrossRef]

48. Cardiovascular Disease: Risk Assessment and Reduction, Including Lipid Modification. National Institute for Health and Care Excellence. 2016. Available online: https://www.nice.org.uk/guidance/cg181 (accessed on 23 July 2021).

49. Collet, J.P.; Thiele, H.; Barbato, E.; Barthélémy, O.; Bauersachs, J.; Bhatt, D.L.; Dendale, P.; Dorobantu, M.; Edvardsen, T.; Folliguet, T.; et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur. Heart J.* **2021**, *42*, 1289–1367. [CrossRef] [PubMed]

50. Huang, W.; FitzGerald, G.; Goldberg, R.J.; Gore, J.; McManus, R.H.; Awad, H.; Waring, M.E.; Allison, J.; Saczynski, J.S.; Kiefe, C.I.; et al. Performance of the GRACE Risk Score 2.0 Simplified Algorithm for Predicting 1-year Death Following Hospitalization for an Acute Coronary Syndrome in a Contemporary Multiracial Cohort. *Am. J. Cardiol.* **2016**, *118*, 1105–1110. [CrossRef] [PubMed]

51. van Rosendael, A.R.; Maliakal, G.; Kolli, K.K.; Beecy, A.; Al'Aref, S.J.; Dwivedi, A.; Singh, G.; Panday, M.; Kumar, A.; Ma, X.; et al. Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the CONFIRM registry. *J. Cardiovasc. Comput. Tomogr.* **2018**, *12*, 204–209. [CrossRef]

52. Al-Shamsi, S.; Govender, R.D. Validation of the REduction of Atherothrombosis for Continued Health (REACH) prediction model for recurrent cardiovascular disease among United Arab Emirates Nationals. *BMC Res. Notes* **2020**, *13*, 484. [CrossRef]

53. Rossello, X.; Dorresteijn, J.A.; Janssen, A.; Lambrinou, E.; Scherrenberg, M.; Bonnefoy-Cudraz, E.; Cobain, M.; Piepoli, M.F.; Visseren, F.L.; Dendale, P.; et al. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardio-vascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *Eur. J. Prev. Cardiol.* **2019**, *26*, 1534–1544. [CrossRef] [PubMed]

54. van't Klooster, C.C.; Bhatt, D.L.; Steg, P.G.; Massaro, J.M.; Dorresteijn, J.A.N.; Westerink, J.; Ruigrok, Y.M.; de Borst, G.J.; Asselbergs, F.W.; van der Graaf, Y.; et al. Predicting 10-year risk of recurrent cardiovascular events andcardiovascular in-terventions in patients with established cardiovascular disease: Results from UCC-SMART and REACH. *Int. J. Cardiol.* **2021**, *325*, 140–148. [CrossRef] [PubMed]

55. Inouye, M.; Abraham, G.; Nelson, C.P.; Wood, A.M.; Sweeting, M.J.; Dudbridge, F.; Lai, F.Y.; Kaptoge, S.; Brozynska, M.; Wang, T.; et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults. *J. Am. Coll. Cardiol.* **2018**, *72*, 1883–1893. [CrossRef] [PubMed]

56. Oemrawsingh, R.M.; Cheng, J.M.; Akkerhuis, K.M.; Kardys, I.; Degertekin, M.; van Geuns, R.J.; Daemen, J.; Boersma, E.; Serruys, P.W.; van Domburg, R.T. High-sensitivity C-reactive protein predicts 10-year cardiovascular outcome after percutaneous coronary intervention. *EuroIntervention* **2016**, *12*, 345–351. [CrossRef] [PubMed]

57. Goncalves, I.; Bengtsson, E.; Colhoun, H.M.; Shore, A.C.; Palombo, C.; Natali, A.; Edsfeldt, A.; Duner, P.; Fredrikson, G.N.; Björkbacka, H.; et al. Elevated Plasma Levels of MMP-12 Are Associated with Atherosclerotic Burden and Symptomatic Cardiovascular Disease in Subjects With Type 2 Diabetes. *Arterioscler. Thromb. Vasc. Biol.* **2015**, *35*, 1723–1731. [CrossRef]

58. Eggers, K.M.; Kempf, T.; Larsson, A.; Lindahl, B.; Venge, P.; Wallentin, L.; Wollert, K.C.; Lind, L. Evaluation of Temporal Changes in Cardiovascular Biomarker Concentrations Improves Risk Prediction in an Elderly Population from the Community. *Clin. Chem.* **2016**, *62*, 485–493. [CrossRef] [PubMed]

59. Rezaee, M.; Putrenko, I.; Takeh, A.; Ganna, A.; Ingelsson, E. Development and validation of risk prediction models for multiple cardiovascular diseases and Type 2 diabetes. *PLoS ONE* **2020**, *15*, e0235758. [CrossRef] [PubMed]

60. Khera, A.V.; Kathiresan, S. Genetics of coronary artery disease: Discovery, biology and clinical translation. *Nat. Rev. Genet.* **2017**, *18*, 331–344. [CrossRef]

61. Hughes, M.F.; Saarela, O.; Stritzke, J.; Kee, F.; Silander, K.; Klopp, N.; Kontto, J.; Karvanen, J.; Willenborg, C.; Salomaa, V.; et al. Genetic Markers Enhance Coronary Risk Prediction in Men: The MORGAM Prospective Cohorts. *PLoS ONE* **2012**, *7*, e40922. [CrossRef]

62. Mega, J.L.; Stitziel, N.O.; Smith, J.G.; Chasman, D.I.; Caulfield, M.J.; Devlin, J.J.; Nordio, F.; Hyde, C.L.; Cannon, C.P.; Sacks, F.M.; et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: An analysis of primary and secondary prevention trials. *Lancet* **2015**, *385*, 2264–2271. [CrossRef]

63. Schunkert, H.; Samani, N.J. Statin treatment: Can genetics sharpen the focus? *Lancet* **2015**, *385*, 2227–2229. [CrossRef]

64. Hall, K.T.; Kessler, T.; Buring, J.E.; Passow, D.; Sesso, H.D.; Zee, R.Y.L.; Ridker, P.M.; Chasman, D.I.; Schunkert, H. Genetic variation at the coronary artery disease risk locus GUCY1A3 modifies cardiovascular disease prevention effects of aspirin. *Eur. Heart J.* **2019**, *40*, 3385–3392. [CrossRef]

65. Marston, N.A.; Kamanu, F.K.; Nordio, F.; Gurmu, Y.; Roselli, C.; Sever, P.S.; Pedersen, T.R.; Keech, A.C.; Wang, H.; Lira Pineda, A.; et al. Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score. *Circulation* **2020**, *141*, 616–623. [CrossRef]

66. Damask, A.; Steg, P.G.; Schwartz, G.G.; Szarek, M.; Hagström, E.; Badimon, L.; Chapman, M.J.; Boileau, C.; Tsimikas, S.; Ginsberg, H.N.; et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. *Circulation* **2020**, *141*, 624–636. [CrossRef]

67. Ehret, G.B.; Ferreira, T.; Chasman, D.I.; Jackson, A.U.; Schmidt, E.M.; Johnson, T.; Thorleifsson, G.; Luan, J.; Donnelly, L.A.; Kanoni, S.; et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* **2016**, *48*, 1171–1184. [CrossRef]

68. Liu, H.H.; Cao, Y.X.; Jin, J.L.; Zhang, H.W.; Hua, Q.; Li, Y.F.; Guo, Y.L.; Zhu, C.G.; Wu, N.Q.; Gao, Y.; et al. Association of lipoprotein(a) levels with recurrent events in patients with coronary artery disease. *Heart* **2020**, *106*, 1228–1235. [CrossRef]

69. Hilvo, M.; Meikle, P.J.; Pedersen, E.R.; Tell, G.S.; Dhar, I.; Brenner, H.; Schöttker, B.; Lääperi, M.; Kauhanen, D.; Koistinen, K.M.; et al. Development and validation of a ceramide- and phospholipid-based cardiovascular risk estimation score for coronary artery disease patients. *Eur. Heart J.* **2020**, *41*, 371–380. [CrossRef]

70. Laaksonen, R.; Ekroos, K.; Sysi-Aho, M.; Hilvo, M.; Vihervaara, T.; Kauhanen, D.; Suoniemi, M.; Hurme, R.; März, W.; Scharnagl, H.; et al. S Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *Eur. Heart J.* **2016**, *37*, 1967–1976. [CrossRef] [PubMed]

71. Patel, R.S.; Schmidt, A.F.; Tragante, V.; McCubrey, R.O.; Holmes, M.V.; Howe Laurence, J.; Kenan, D.; Axel, Å.; Karin, L.; Virani Salim, S.; et al. Association of Chromosome 9p21 With Subsequent Coronary Heart Disease Events. *Circ. Genom. Precis. Med.* **2019**, *12*, e002471. [CrossRef] [PubMed]

72. Mahmoodi, B.K.; Tragante, V.; Kleber Marcus, E.; Holmes, M.V.; Schmidt, A.F.; McCubrey Raymond, O.; Howe Laurence, J.; Kenan, D.; Hooman, A.; Baranova Ekaterina, V.; et al. Association of Factor V Leiden With Subsequent Atherothrombotic Events. *Circulation* **2020**, *142*, 546–555. [CrossRef] [PubMed]

73. Helgadottir, A.; Thorleifsson, G.; Manolescu, A.; Gretarsdottir, S.; Blondal, T.; Jonasdottir, A.; Jonasdottir, A.; Sigurdsson, A.; Baker, A.; Palsson, A.; et al. A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. *Science* **2007**, *316*, 1491–1493. [CrossRef]

74. Talukdar, H.A.; Asl, H.F.; Jain, R.K.; Ermel, R.; Ruusalepp, A.; Franzén, O.; Kidd, B.A.; Readhead, B.; Giannarelli, C.; Kovacic, J.C.; et al. Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. *Cell Syst.* **2016**, *2*, 196–208. [CrossRef] [PubMed]

75. Lempiäinen, H.; Brænne, I.; Michoel, T.; Tragante, V.; Vilne, B.; Webb, T.R.; Kyriakou, T.; Eichner, J.; Zeng, L.; Willenborg, C.; et al. Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. *Sci. Rep.* **2018**, *8*, 3434. [CrossRef] [PubMed]

76. Foroughi Asl, H.; Talukdar, H.A.; Kindt, A.S.; Jain, R.K.; Ermel, R.; Ruusalepp, A.; Nguyen, K.D.H.; Dobrin, R.; Reilly, D.F.; Schunkert, H.; et al. Expression Quantitative Trait Loci Acting Across Multiple Tissues Are Enriched in Inherited Risk for Coronary Artery Disease. *Circ. Cardiovasc. Genet.* **2015**, *8*, 305–315. [CrossRef] [PubMed]

77. Sun, L.; Pennells, L.; Kaptoge, S.; Nelson, C.P.; Ritchie, S.C.; Abraham, G.; Arnold, M.; Bell, S.; Bolton, T.; Burgess, S.; et al. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. *PLoS Med.* **2021**, *18*, e1003498. [CrossRef]

78. Levin, M.G.; Rader, D.J. Polygenic Risk Scores and Coronary Artery Disease. *Circulation* **2020**, *141*, 637–640. [CrossRef]

79. Howe, L.J.; Dudbridge, F.; Schmidt, A.F.; Finan, C.; Denaxas, S.; Asselbergs, F.W.; Hingorani, A.D.; Patel, R.S. Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Hum. Mol. Genet.* **2020**, *29*, 1388–1395. [CrossRef]

80. Qian, J.; Tanigawa, Y.; Du, W.; Aguirre, M.; Chang, C.; Tibshirani, R.; Rivas, M.A.; Hastie, T. A fast and scalable framework for large-scale and ultrahigh-dimensional sparse regression with application to the UK Biobank. *PLoS Genet.* **2020**, *16*, e1009141. [CrossRef] [PubMed]

81. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L.T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O'Connell, J.; et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**, *562*, 203–209. [CrossRef]

82. Elliott, J.; Bodinier, B.; Bond, T.A.; Chadeau-Hyam, M.; Evangelou, E.; Moons, K.G.M.; Dehghan, A.; Muller, D.C.; Elliott, P.; Tzoulaki, I. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs. a Clinical Risk Score for Coronary Artery Disease. *J. Am. Med. Assoc.* **2020**, *323*, 636–645. [CrossRef]

83. Bauer, A.; Zierer, A.; Gieger, C.; Büyüközkan, M.; Müller-Nurasyid, M.; Grallert, H.; Meisinger, C.; Strauch, K.; Prokisch, H.; Roden, M.; et al. Comparison of genetic risk prediction models to improve prediction of coronary heart disease in two large cohorts of the MONICA/KORA study. *Genet. Epidemiol.* **2021**, *45*, 633–650. [CrossRef]

84. Weijmans, M.; de Bakker, P.I.W.; van der Graaf, Y.; Asselbergs, F.W.; Algra, A.; de Borst, G.J.; Spiering, W.; Visseren, F.L.J. Incremental value of a genetic risk score for the prediction of new vascular events in patients with clinically manifest vascular disease. *Atherosclerosis* **2015**, *239*, 451–458. [CrossRef] [PubMed]

85. Jiang, J.; Zheng, Q.; Han, Y.; Qiao, S.; Chen, J.; Yuan, Z.; Yu, B.; Ge, L.; Jia, J.; Gong, Y.; et al. Genetic predisposition to coronary artery disease is predictive of recurrent events: A Chinese prospective cohort study. *Hum. Mol. Genet.* **2020**, *29*, 1044–1053. [CrossRef] [PubMed]

86. Herrmann, M.; Probst, P.; Hornung, R.; Jurinovic, V.; Boulesteix, A.L. Large-scale benchmark study of survival prediction methods using multi-omics data. *Briefings Bioinform.* **2021**, *22*, bbaa167. [CrossRef] [PubMed]

87. Blanco-Colio, L.M.; Mendez-Barbero, N.; Pello Lazaro, A.M.; Acena, A.; Tarin, N.; Cristobal, C.; Martinez-Milla, J.; Gonzalez-Lorenzo, O.; Martin-Ventura, J.L.; Huelmos, A.; et al. MCP-1 Predicts Recurrent Cardiovascular Events in Patients with Persistent Inflammation. *J. Clin. Med.* **2021**, *10*, 1137. [CrossRef]

88. Tunon, J.; Blanco-Colio, L.; Cristóbal, C.; Tarín, N.; Higueras, J.; Huelmos, A.; Alonso, J.; Egido, J.; Asensio, D.; Lorenzo, O.; et al. Usefulness of a combination of monocyte chemoattractant protein-1, galectin-3, and N-terminal probrain natriuretic peptide to predict cardiovascular events in patients with coronary artery disease. *Am. J. Cardiol.* **2014**, *113*, 434–440. [CrossRef]

89. Nguyen, N.D.; Wang, D. Multiview learning for understanding functional multiomics. *PLoS Comput. Biol.* **2020**, *16*, e1007677. [CrossRef]

90. Miller, C.L.; Pjanic, M.; Wang, T.; Nguyen, T.; Cohain, A.; Lee, J.D.; Perisic, L.; Hedin, U.; Kundu, R.K.; Majmudar, D.; et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* **2016**, *7*, 12092. [CrossRef]

91. Barbeira, A.N.; Bonazzola, R.; Gamazon, E.R.; Liang, Y.; Park, Y.; Kim-Hellmuth, S.; Wang, G.; Jiang, Z.; Zhou, D.; Hormozdiari, F.; et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **2021**, *22*, 49. [CrossRef]

92. Huan, T.; Zhang, B.; Wang, Z.; Joehanes, R.; Zhu, J.; Johnson, A.D.; Ying, S.; Munson, P.J.; Raghavachari, N.; Wang, R.; et al. A Systems Biology Framework Identifies Molecular Underpinnings of Coronary Heart Disease. *Arterioscler. Thromb. Vasc. Biol.* **2013**, *33*, 1427–1434. [CrossRef]

93. Heinig, M.; Adriaens, M.E.; Schafer, S.; van Deutekom, H.W.M.; Lodder, E.M.; Ware, J.S.; Schneider, V.; Felkin, L.E.; Creemers, E.E.; Meder, B.; et al. Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol.* **2017**, *18*, 170. [CrossRef]

94. Gene Ontology Consortium. Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **2015**, *43*, D1049–D1056. [CrossRef]

95. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]

96. Yao, F.; Zhang, K.; Zhang, Y.; Guo, Y.; Li, A.; Xiao, S.; Liu, Q.; Shen, L.; Ni, J. Identification of Blood Biomarkers for Alzheimer's Disease Through Computational Prediction and Experimental Validation. *Front. Neurol.* **2019**, *9*, 1158. [CrossRef]

97. Kanuri, S.H.; Ipe, J.; Kassab, K.; Gao, H.; Liu, Y.; Skaar, T.C.; Kreutz, R.P. Next generation MicroRNA sequencing to identify coronary artery disease patients at risk of recurrent myocardial infarction. *Atherosclerosis* **2018**, *278*, 232–239. [CrossRef]

98. Ghatge, M.; Sharma, A.; Maity, S.; Kakkar, V.V.; Vangala, R.K. Danger-recognizing proteins, beta-defensin-128 and histatin-3, as potential biomarkers of recurrent coronary events. *Int. J. Mol. Med.* **2017**, *40*, 531–538. [CrossRef]

99. Seifert, S.; Gundlach, S.; Junge, O.; Szymczak, S. Integrating biological knowledge and gene expression data using pathway-guided random forests: A benchmarking study. *Bioinformatics* **2020**, *36*, 4301–4308. [CrossRef] [PubMed]

100. Nazarieh, M.; Rajula, H.S.R.; Helms, V. Topology Consistency of Disease-specific Differential Co-regulatory Networks. *BMC Bioinform.* **2019**, *20*, 550. [CrossRef] [PubMed]

101. Zuo, Y.; Cui, Y.; Yu, G.; Li, R.; Ressom, H.W. Incorporating prior biological knowledge for network-based differential gene expression analysis using differentially weighted graphical LASSO. *BMC Bioinform.* **2017**, *18*, 99. [CrossRef] [PubMed]

102. Wang, L.; Hu, J.; Zhou, J.; Guo, F.; Yao, T.; Zhang, L. Weighed Gene Coexpression Network Analysis Screens the Potential Long Noncoding RNAs and Genes Associated with Progression of Coronary Artery Disease. *Comput. Math. Methods Med.* **2020**, *2020*, e8183420, [CrossRef] [PubMed]

103. Franzén, O.; Ermel, R.; Cohain, A.; Akers, N.K.; Narzo, A.D.; Talukdar, H.A.; Foroughi-Asl, H.; Giambartolomei, C.; Fullard, J.F.; Sukhavasi, K.; et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **2016**, *353*, 827–830. [CrossRef]

104. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97. [CrossRef] [PubMed]

105. Tragante, V.; Hemerich, D.; Alshabeeb, M.; Brænne, I.; Lempiäinen, H.; Patel, R.S.; den Ruijter, H.M.; Barnes, M.R.; Moore, J.H.; Schunkert, H.; et al. Druggability of Coronary Artery Disease Risk Loci. *Circ. Genom. Precis. Med.* **2018**, *11*, e001977. [CrossRef]

106. Zeng, L.; Talukdar, H.A.; Koplev, S.; Giannarelli, C.; Ivert, T.; Gan, L.M.; Ruusalepp, A.; Schadt, E.E.; Kovacic, J.C.; Lusis, A.J.; et al. Contribution of Gene Regulatory Networks to Heritability of Coronary Artery Disease. *J. Am. Coll. Cardiol.* **2019**, *73*, 2946–2957. [CrossRef] [PubMed]

107. Sumathipala, M.; Weiss, S.T. Predicting miRNA-based disease-disease relationships through network diffusion on multi-omics biological data. *Sci. Rep.* **2020**, *10*, 8705. [CrossRef] [PubMed]

108. Chou, C.H.; Shrestha, S.; Yang, C.D.; Chang, N.W.; Lin, Y.L.; Liao, K.W.; Huang, W.C.; Sun, T.H.; Tu, S.J.; Lee, W.H.; et al. miRTar-Base update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **2018**, *46*, D296–D302. [CrossRef] [PubMed]

109. Hamosh, A.; Scott, A.F.; Amberger, J.; Valle, D.; McKusick, V.A. Online Mendelian Inheritance In Man (OMIM). *Hum. Mutat.* **2000**, *15*, 57–61. [CrossRef]

110. Piñero, J.; Ramírez-Anguita, J.M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; Furlong, L.I. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **2020**, *48*, D845–D855. [CrossRef] [PubMed]

111. Ghafouri-Fard, S.; Gholipour, M.; Taheri, M. Role of MicroRNAs in the Pathogenesis of Coronary Artery Disease. *Front. Cardiovasc. Med.* **2021**, *8*, 632392. [CrossRef]

112. Oughtred, R.; Rust, J.; Chang, C.; Breitkreutz, B.; Stark, C.; Willems, A.; Boucher, L.; Leung, G.; Kolas, N.; Zhang, F.; et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci. Publ. Protein Soc.* **2021**, *30*, 187–200. [CrossRef]

113. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef]

114. Kamburov, A.; Pentchev, K.; Galicka, H.; Wierling, C.; Lehrach, H.; Herwig, R. ConsensusPathDB: Toward a more complete picture of cell biology. *Nucleic Acids Res.* **2011**, *39*, D712–D717. [CrossRef]

115. Bader, G.D.; Cary, M.P.; Sander, C. Pathguide: A Pathway Resource List. *Nucleic Acids Res.* **2006**, *34*, D504–D506. [CrossRef]

116. Ghatge, M.; Nair, J.; Sharma, A.; Vangala, R.K. Integrative gene ontology and network analysis of coronary artery disease associated genes suggests potential role of ErbB pathway gene EGFR. *Mol. Med. Rep.* **2018**, *17*, 4253–4264. [CrossRef]

117. Du, P.; Feng, G.; Flatow, J.; Song, J.; Holko, M.; Kibbe, W.A.; Lin, S.M. From disease ontology to disease-ontology lite: Statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations. *Bioinformatics* **2009**, *25*, i63–i68. [CrossRef]

118. Liu, H.; Liu, W.; Liao, Y.; Cheng, L.; Liu, Q.; Ren, X.; Shi, L.; Tu, X.; Wang, Q.K.; Guo, A.Y. CADgene: A comprehensive database for coronary artery disease genes. *Nucleic Acids Res.* **2011**, *39*, D991–D996. [CrossRef]

119. Kazeminia, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; Mukhopadhyay, A. GANs for medical image analysis. *Artif. Intell. Med.* **2020**, *109*, 101938. [CrossRef]

120. Christiansen, F.; Epstein, E.L.; Smedberg, E.; Åkerlund, M.; Smith, K.; Epstein, E. Ultrasound image analysis using deep neural networks for discriminating between benign and malignant ovarian tumors: Comparison with expert subjective assessment. *Ultrasound Obstet. Gynecol.* **2021**, *57*, 155–163. [CrossRef] [PubMed]

121. Matek, C.; Schwarz, S.; Spiekermann, K.; Marr, C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat. Mach. Intell.* **2019**, *1*, 538–544. [CrossRef]

122. Schouten, J.P.E.; Matek, C.; Jacobs, L.F.P.; Buck, M.C.; Bošnački, D.; Marr, C. Tens of images can suffice to train neural networks for malignant leukocyte detection. *Sci. Rep.* **2021**, *11*, 7995. [CrossRef] [PubMed]

123. Bai, W.; Suzuki, H.; Huang, J.; Francis, C.; Wang, S.; Tarroni, G.; Guitton, F.; Aung, N.; Fung, K.; Petersen, S.E.; et al. A population-based phenome-wide association study of cardiac and aortic structure and function. *Nat. Med.* **2020**, *26*, 1654–1662. [CrossRef]

124. Warnat-Herresthal, S.; Schultze, H.; Shastry, K.L.; Manamohan, S.; Mukherjee, S.; Garg, V.; Sarveswara, R.; Händler, K.; Pickkers, P.; Aziz, N.A.; et al. Swarm Learning for decentralized and confidential clinical machine learning. *Nature* **2021**, *594*, 265–270. [CrossRef] [PubMed]

125. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef] [PubMed]

126. Ward, A.; Sarraju, A.; Chung, S.; Li, J.; Harrington, R.; Heidenreich, P.; Palaniappan, L.; Scheinker, D.; Rodriguez, F. Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population. *NPJ Digit. Med.* **2020**, *3*, 125. [CrossRef] [PubMed]

127. Junwei, K.; Yang, H.; Junjiang, L.; Zhijun, Y. Dynamic prediction of cardiovascular disease using improved LSTM. *Int. J. Crowd Sci.* **2019**, *3*, 14–25. [CrossRef]

128. Hossain, M.E.; Uddin, S.; Khan, A. Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. *Expert Syst. Appl.* **2021**, *164*, 113918. [CrossRef]

129. D'Ascenzo, F.; Filippo, O.D.; Gallone, G.; Mittone, G.; Deriu, M.A.; Iannaccone, M.; Ariza-Solé, A.; Liebetrau, C.; Manzano-Fernández, S.; Quadri, G.; et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): A modelling study of pooled datasets. *Lancet* **2021**, *397*, 199–207. [CrossRef]

130. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995 ; Volume 1, pp. 278–282. [CrossRef]

131. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [CrossRef]

132. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

133. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]

134. Bai, W.; Sinclair, M.; Tarroni, G.; Oktay, O.; Rajchl, M.; Vaillant, G.; Lee, A.M.; Aung, N.; Lukaschuk, E.; Sanghvi, M.M.; et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **2018**, *20*, 65. [CrossRef]

135. Wolterink, J.M.; Leiner, T.; de Vos, B.D.; Coatrieux, J.L.; Kelm, B.M.; Kondo, S.; Salgado, R.A.; Shahzad, R.; Shu, H.; Snoeren, M.; et al. An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework. *Med. Phys.* **2016**, *43*, 2361. [CrossRef]

136. Litjens, G.; Ciompi, F.; Wolterink, J.M.; de Vos, B.D.; Leiner, T.; Teuwen, J.; Išgum, I. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC Cardiovasc. Imaging* **2019**, *12*, 1549–1565. [CrossRef]

137. Zhou, J.; Troyanskaya, O.G. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods* **2015**, *12*, 931–934. [CrossRef]

138. Talukder, A.; Barham, C.; Li, X.; Hu, H. Interpretation of deep learning in genomics and epigenomics. *Briefings Bioinform.* **2021**, *22*, bbaa177. [CrossRef]

139. Eraslan, G.; Avsec, Z.; Gagneur, J.; Theis, F.J. Deep learning: New computational modelling techniques for genomics. *Nat. Rev. Genet.* **2019**, *20*, 389–403. [CrossRef] [PubMed]

140. Yin, B.; Balvert, M.; van der Spek, R.A.A.; Dutilh, B.E.; Bohté, S.; Veldink, J.; Schönhuth, A. Using the structure of genome data in the design of deep neural networks for predicting amyotrophic lateral sclerosis from genotype. *Bioinformatics* **2019**, *35*, i538–i547. [CrossRef] [PubMed]

141. Dutta, A.; Batabyal, T.; Basu, M.; Acton, S.T. An efficient convolutional neural network for coronary heart disease prediction. *Expert Syst. Appl.* **2020**, *159*, 113408. [CrossRef]

142. Linden, A.; Yarnold, P.R. Modeling time-to-event (survival) data using classification tree analysis. *J. Eval. Clin. Pract.* **2017**, *23*, 1299–1308. [CrossRef]

143. Chen, X.; Ishwaran, H. Pathway hunting by random survival forests. *Bioinformatics* **2013**, *29*, 99–105. [CrossRef] [PubMed]

144. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **2018**, *18*, 24. [CrossRef]

145. Kvamme, H.; Borgan, O.; Scheel, I. Time-to-Event Prediction with Neural Networks and Cox Regression. *J. Mach. Learn. Res.* **2019**, *20*, 1–30.

146. Schäfer, J.; Strimmer, K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **2005**, *21*, 754–764. [CrossRef]

147. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **2008**, *9*, 432–441. [CrossRef] [PubMed]

148. Hawe, J.S.; Theis, F.J.; Heinig, M. Inferring Interaction Networks From Multi-Omics Data. *Front. Genet.* **2019**, *10*, 535. [CrossRef] [PubMed]

149. Zierer, J.; Pallister, T.; Tsai, P.C.; Krumsiek, J.; Bell, J.T.; Lauc, G.; Spector, T.D.; Menni, C.; Kastenmüller, G. Exploring the molecular basis of age-related disease comorbidities using a multi-omics graphical model. *Sci. Rep.* **2016**, *6*, 37646. [CrossRef] [PubMed]

150. Manatakis, D.V.; Raghu, V.K.; Benos, P.V. piMGM: Incorporating multi-source priors in mixed graphical models for learning disease networks. *Bioinformatics* **2018**, *34*, i848–i856. [CrossRef] [PubMed]

151. Zhang, X.M.; Liang, L.; Liu, L.; Tang, M.J. Graph Neural Networks and Their Current Applications in Bioinformatics. *Front. Genet.* **2021**, *12*, 690049. [CrossRef] [PubMed]

152. Wang, T.; Shao, W.; Huang, Z.; Tang, H.; Zhang, J.; Ding, Z.; Huang, K. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat. Commun.* **2021**, *12*, 3445. [CrossRef]

153. Schulte-Sasse, R.; Budach, S.; Hnisz, D.; Marsico, A. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nat. Mach. Intell.* **2021**, *3*, 513–526. [CrossRef]

154. Chereda, H.; Bleckmann, A.; Menck, K.; Perera-Bel, J.; Stegmaier, P.; Auer, F.; Kramer, F.; Leha, A.; Beißbarth, T. Explaining decisions of graph convolutional neural networks: Patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer. *Genome Med.* **2021**, *13*, 42. [CrossRef]

155. Rhee, S.; Seo, S.; Kim, S. Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification. *arXiv* **2018**, arXiv:1711.05859.

156. Young, J.; Modat, M.; Cardoso, M.J.; Mendelson, A.; Cash, D.; Ourselin, S. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin.* **2013**, *2*, 735–745. [CrossRef]

157. DeGrave, A.J.; Janizek, J.D.; Lee, S.I. AI for radiographic COVID-19 detection selects shortcuts over signal. *Nat. Mach. Intell.* **2021**, *3*, 610–619. [CrossRef]

158. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.R. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. *Proc. IEEE* **2021**, *109*, 247–278. [CrossRef]

159. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

160. Fleetwood, O.; Kasimova, M.A.; Westerlund, A.M.; Delemotte, L. Molecular Insights from Conformational Ensembles via Machine Learning. *Biophys. J.* **2020**, *118*, 765–780. [CrossRef] [PubMed]

161. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1135–1144. [CrossRef]

162. Montavon, G.; Lapuschkin, S.; Binder, A.; Samek, W.; Müller, K.R. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognit.* **2017**, *65*, 211–222. [CrossRef]

163. Eichler, G.S.; Reimers, M.; Kane, D.; Weinstein, J.N. The LeFE algorithm: Embracing the complexity of gene expression in the interpretation of microarray data. *Genome Biol.* **2007**, *8*, R187. [CrossRef]

164. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]

165. Hausken, K.; Mohr, M. The value of a player in n-person games. *Soc. Choice Welf.* **2001**, *18*, 465–483. [CrossRef]

166. Eitel, F.; Soehler, E.; Bellmann-Strobl, J.; Brandt, A.U.; Ruprecht, K.; Giess, R.M.; Kuchling, J.; Asseyer, S.; Weygandt, M.; Haynes, J.D.; et al. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *Neuroimage Clin.* **2019**, *24*, 102003. [CrossRef] [PubMed]

167. Böhle, M.; Eitel, F.; Weygandt, M.; Ritter, K. Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer's Disease Classification. *Front. Aging Neurosci.* **2019**, *11*, 194. [CrossRef] [PubMed]